2018

Social interactions and the prisoner's dilemma game: new measures of cognitive and behavioral phenotypes

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

SOCIAL INTERACTIONS AND THE PRISONER'S DILEMMA GAME: NEW MEASURES OF COGNITIVE AND BEHAVIORAL PHENOTYPES

by

MEIA CHITA-TEGMARK

B.A., Babes-Bolyai University, Cluj-Napoca, Romania, 2008
M.Ed., Harvard University, 2010

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

2018

Approved by

First Reader

_____

Peter R. Blake, Ed.D.
Assistant Professor of Psychological and Brain Sciences
Boston University

Second Reader

_____

Kimberly Saudino, Ph.D.
Professor of Psychological and Brain Sciences
Boston University

Third Reader

_____

Melissa Kibbe, Ph.D.
Assistant Professor of Psychological and Brain Sciences
Boston University

Fourth Reader

_____

Kristin Long, Ph.D.
Assistant Professor of Psychological and Brain Sciences
Boston University

ACKNOWLEDGEMENTS

I am truly grateful to everyone who has offered me support and guidance in completing this dissertation, including

my advisor Peter Blake for taking a chance on me, and for showing me not just how to do good research but also what it means to be a perceptive, caring and devoted mentor,

the SDLL lab, for being my second family,

my amazing colleagues and friends Nadia Chernyak, Telli Davoodi and Tess Harvey, for their endless warmth and inspiration,

my research assistant and friend Claire Park for her helpful hard work on this research project,

the teachers and students at Acera and Banneker Schools for making me nostalgic about being in a classroom again,

the children and parents who took precious time to complete this online study,

the professors who have taught and advised me along the way, Sudha Arunachalam, Deborah Kelemen, Melissa Kibbe, Kristin Long, Charles Nelson, Kimberly Saudino, Helen Tager-Flusberg, Amanda Tarullo, and many others,

my parents for their unwavering support of me,

my sister Daniela for all her encouragement and love,

my brother George for keeping me forever curious about the minds of children,

and most of all, my beloved husband, Max, for always believing in me, for seeing more in me than I was ever able to, and for never letting me give up.

Finally, I'd like to thank the child in my study who explained why and how he played my game in the following way: "because GO BIG OR GO HOME! And I never EVER lose!" – it's become the humorous mantra that has carried me through all the moments of doubt and exhaustion in completing this dissertation.

SOCIAL INTERACTIONS AND THE PRISONER'S DILEMMA GAME: NEW MEASURES OF COGNITIVE AND BEHAVIORAL PHENOTYPES

MEIA CHITA-TEGMARK

Boston University Graduate School of Arts and Sciences, 2018

Major Professor: Peter R. Blake, Assistant Professor of Psychological and Brain Sciences

ABSTRACT

Healthy social interactions are critical for children's development and academic and life success. Relevant research is found in two key disciplines with different emphases: Developmental psychology focuses on individual thoughts, motivations and traits; and behavioral economics and game theory focuses on behavioral tasks. This project integrated these approaches by validating a game-theoretic task for children, the Repeated Prisoner's Dilemma (RPD), and demonstrated how it can be used to elucidate the mechanisms underlying children's social interactions.

I developed a novel RPD with fixed-strategy partners in order to test specific hypotheses based on developmental theories of social interaction. Children between 9 and 11 years of age ($N = 167$) were tested on the RPD followed by questions about how they played and interpreted the task. Parents completed a questionnaire assessing

their child's reactive and proactive aggressive traits, a basis for predicting decisions in the RPD. Children also completed a Social Information Processing (SIP) task with novel positive scenarios in addition to standard negative ones.

I hypothesized that: 1) children would interpret the RPD as a real social interaction and engage in strategic forms of play according to game theory; 2) children with different levels of reactive and proactive aggression would show different patterns of RPD play based on theories of aggression; 3) the SIP responses would predict different levels of cooperation in the RPD, and the positive scenarios would generate responses consistent with the general SIP theory.

Results showed support for the first hypothesis with classes of motivations (interpersonal and strategic self-interest) predicting RPD behavior. The second hypothesis was partially confirmed: Children rated high on reactive aggression showed reactive responses in the RPD. This analysis also revealed an important novel finding that high-reactive children followed a game-theoretic strategy known as "Grim" – they did not return to cooperation after partner defection. The third hypothesis was partially confirmed: Responses for the positive scenarios were consistent with the SIP model but did not predict RPD play.

These findings demonstrate the value of integrating theoretical and methodological approaches from developmental psychology and game theory in order to study the mechanisms of social interaction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND BACKGROUND: INTEGRATING APPROACHES FROM DEVELOPMENTAL PSYCHOLOGY AND GAME THEORY TO STUDY SOCIAL INTERACTIONS

Healthy social interactions are critical for individual wellbeing, as well as academic and life success (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). As such, the study of both positive and negative social interactions has been central to understanding children's development and improving their life outcomes. However, within developmental psychology, the study of positive and negative social behaviors has evolved separately to a large extent. For example, well-developed theories and taxonomies of aggressive behavior have remained detached from developmental theories of prosocial behavior. The most well-established model for understanding the psychology of social interactions, the Social Information Processing (SIP) model, has focused primarily on negative interactions despite being conceived as a general theory of both positive and negative interactions. Methodological approaches are also very different with negative behaviors mostly studied using hypothetical vignettes and questionnaires and positive behaviors studied with a combination of questionnaires and behavioral tasks.

In contrast to this divide within developmental psychology, positive and negative social interactions have been studied together by researchers utilizing the tools of

behavioral economics and game theory. Behavioral economists, evolutionary theorists and psychologists have long used interactive tasks with real rewards to understand human decisions. These so-called "games" have revealed much about how cooperation emerges and stabilizes in societies given the tension between positive motivations (altruism, trust) and negative or strategic motivations (punishment, self-interest). Despite these advances, much less is known about the psychological mechanisms that underlie social interactions within the context of games and how these decisions align with psychological traits and models of social cognition. Moreover, developmental psychologists have only begun to adopt these tools in recent years and the links between children's decisions in games and developmental theories remain unclear.

This dissertation represents an attempt to bridge this theoretical and methodological divide by combining a behavioral measure from game theory, the Repeated Prisoner's Dilemma (RPD), with more traditional questionnaires and vignettes from developmental psychology. In what follows, we review theory and evidence for two forms of aggression – reactive and proactive – and highlight the need for new behavioral tasks. We then review the SIP model and point out that even though it has been extensively tested for negative interactions, it is unclear whether the same mechanisms apply universally, also for positive social interactions. Finally, we review the Prisoner's Dilemma (PD) game and its use with children, arguing that it can serve as a safe, quantitatively rigorous task for measuring behavioral phenotypes in both negative and positive social interactions.

Reactive and proactive aggression

One of the more well-developed lines of research on social interactions in children has focused on aggressive behaviors, or when and how social interactions result in hurt, harm or injury. Aggressive actions can be either physical (i.e.,, hitting) or relational (i.e.,, gossip and social exclusion). Aggression is one of the most common behaviors in negative interactions among children. Within the US, bullying (physical or relational) constitutes a major problem in schools. According to a 2009 survey of 6500 children from 6-12 grade, 39.4% experienced bullying in 6th grade, a percentage which decreased to 20.4% by 12 grade (National Center for Education Statistics). Aggression has negative consequences for both aggressors and victims. For aggressors, it can indicate the presence of clinical disorders such as Oppositional Defiant Disorder (ODD), Conduct Disorder (CD) or Conduct Problems (CP), with outcomes sometimes including juvenile delinquency and crime. For both aggressors and victims, negative interactions involving aggressive behavior can lead to internalizing disorders at later times.

Work in developmental psychology has identified two different forms of aggression, reactive and proactive, which differ in the kinds of social interactions that prompt and sustain them (Vitaro, Brendgen, & Baker, 2006). Reactive aggression is characterized by an impulsive defensive response to perceived hostility (Fontaine & Dodge, 2006), while proactive aggression is characterized by controlled aggressive behavior that anticipates a reward (Matthys & Lochman, 2011).

Reactive and proactive aggression are different in several aspects including the trigger of the response, the goal, the time-scale of the response, and physiological

reactions (Kempes, Matthys, de Vries, & van Engeland, 2005). Reactive aggression is triggered by a threat or provocation, which may be either real or perceived. Indeed, one key feature of people with reactive traits is a tendency to perceive ambiguous interactions as intentional and hostile – the so-called hostile attribution bias (Dodge & Coie, 1987). This trigger gives rise to anger and the response is often intended to harm the perceived provocateur, although this may actually be a self-protective or defensive move intended to prevent further provocation. Reactive aggression is enacted on a short time scale and is accompanied by high levels of physiological arousal (Kempes et al., 2005).

Proactive aggression, in contrast, is triggered by the opportunity to obtain a desired material gain. It is characterized by instrumental goals: gaining a benefit from the interaction, for example, by using blackmail or intimidation through physical aggression to obtain desired goods. Proactive aggression is characterized by long-term, calculated behaviors and is associated with low physiological arousal.

The two types of aggression are also distinguished based on differences in developmental origins, trajectory and outcomes: reactive aggression is associated with histories of physical abuse, and is thought to emerge as a reaction to a hostile environment; proactive aggression is not linked to abuse and is believed to emerge in environments that support the use of aggression as a means of achieving one's goals (Dodge, Lochman, Harnish, Bates, & Pettit, 1997). In a review of the literature on developmental trajectories of aggressive behavior Vitaro et al. (2006) note a reported decline in physical aggression between the ages of 4-11, but an increase in social aggression, a more indirect, covert and "circuitous" type of aggression. The authors also

suggest that the increase in self-regulatory capabilities may foster a decline in reactive aggression, while proactive aggression could remain stable or increase during adolescence in environments supportive of aggressive solutions to problem solving and access to resources. Kempes et al. (2005) also point out the physical and impulsive nature of the aggressive behavior in young children compared to the more planned and calculated aggression that appears with age and with improvements in cognitive abilities such as planning. They note that this distinction is partly supported by the finding of Connor, Steingard, Cunningham, Anderson and Melloni (2004) that reactive aggression is negatively related to age but proactive aggression is not and appears to remain stable with age at least for some individuals. This lack of a decline with age may help to explain why proactive aggression, but not reactive aggression, is associated with juvenile delinquency, criminal behavior and ODD and CD diagnoses in mid-adolescence.

Lastly, separate temperamental and possibly genetic factors have been linked to the emergence of the two different types of aggressive behavior. For example, reactive, but not proactive aggression has been linked to anxiety, angry reactivity, emotional dysregulation and inattention (Dodge & Coie, 1987; Dodge et al., 1997; Vitaro et al., 2006). In contrast, proactive, but not reactive, aggression has been linked to callous unemotional traits (Helseth, Waschbusch, King & Willoughby, 2015).

The Social Information Processing model

The SIP model has been proposed as a series of cognitive steps for appraising social interactions, which explain how and why aggressive behaviors are formulated, as

opposed to viewing aggressiveness as simply a trait. The SIP model posits that behaviors are generated through five information processing steps (that can occur in parallel): 1) encoding situational cues, 2) attributing intent, 3) selecting goals, 4) generating possible behavioral responses, and 5) evaluating these responses and selecting one to be enacted (Crick & Dodge, 1994).

Encoding situational cues refers to the number and type of details that are encoded about the situation (Dodge & Newman, 1981), followed by attributions of other people's intentions when they have caused a particular outcome. Thus, if someone bumps into you while walking, one will assess whether this was done on purpose or by accident (Dodge, 2008). Goal selection follows and falls into two general categories: relational goals, which are concerned with enhancing or preserving the relationship (e.g.,, avoiding unnecessary conflict), or instrumental goals, which seek to profit from the interaction (e.g.,, obtaining material gain or asserting dominance; Camodeca & Goossens, 2005; van Rest et al., 2014). In line with the goals adopted, a number of possible responses are formulated that can be either aggressive or socially competent (Dodge & Godwin, 2013). Those responses are then evaluated based on several criteria including how well they would achieve the goal, how competent one is at enacting the response (e.g.,, how physically or emotionally capable one is to respond aggressively) and how socially and morally acceptable the response is (Kupersmidt, Stelter, & Dodge, 2011). Based on the evaluation step a response is selected and enacted.

The SIP model has been successfully tested with regards to the formulation of aggressive behaviors by drawing connections between the steps and the presence and

frequency of aggressive behaviors in daily life, effects that have been found consistent across different ages and cultures (e.g., Dodge et al., 2015). Some studies have also found that reactive and proactive aggression are differentially linked to the SIP model. The propensity to attribute hostile (as opposed to benign) intent to others in ambiguous situations has been linked to the emergence of reactive aggressive behaviors (Crick & Dodge, 1996; de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002). On the other hand, adopting primarily instrumental goals rather than relational ones and evaluating aggression more positively have been linked to proactive aggression (Crick & Dodge, 1996).

Although the SIP has been conceived as a general model of social interactions, much less attention has been given to positive social behaviors. Only a few studies suggest that links can be drawn between the SIP model and prosocial behavior (Andrade et al., 2012; Laible, McGinley, Carlo, Augustine, & Murphy, 2014; Laible, Murphy, & Augustine, 2014; Nelson & Crick, 1999;). Some studies have linked socially competent SIP in hypothetical negative scenarios with prosocial behaviors in real life. For example, Nelson and Crick (1999) found that in negative interactions including provocation, prosocial young adolescents were less likely to attribute hostile intent, evaluated aggression more negatively and were more likely to adopt relational rather than instrumental goals. Other studies have also found longitudinal relations between SIP in hypothetical negative interactions and both aggressive and prosocial real-life behaviors (Laible, Murphy, & Augustine, 2014; Laible, McGinley, Carlo, Augustine, & Murphy 2014).

None of these studies however have assessed the SIP model in positive interactions. It is thus unclear if the same steps in the thought process apply when positive behaviors, such as positive reciprocity, are generated in positive interactions. Andrade et al. (2012) varied the outcome and the intent of the interaction partner in the vignette to include both positive outcomes and ambiguous, positive or negative intentions. The study found differences in SIP between children with ADHD and controls in the types of intent they attributed and the behavioral responses they formulated, but the study did not investigate whether the same links between the SIP steps (intent attribution, goal selection on response evaluation on one hand and response generation on the other) apply in the case of positive outcome interactions. For example, does attributing positive intent predict positive reciprocity?

To have a unified framework for how both negative and positive behaviors occur, mechanisms proposed by the SIP model still need to be tested for positive interactions using positive-outcome vignettes. In addition, the SIP model should be tested along with tasks measuring children's actual positive and negative behavior in order to determine whether the steps of thinking for vignettes predicts decisions in real social interactions.

Questionnaire and vignette-based measures of aggression

Reactive and proactive aggression have been primarily measured though self, parent or teacher report, and the SIP steps associated with aggressive behaviors have primarily been assessed using hypothetical vignettes. Report measures consist of questionnaires listing behaviors that are characteristic of each of the aggression types,

such as “this child strikes back when teased” for reactive aggression or “this child uses force to dominate peers” for proactive aggression (Kempes et al., 2005). The first such questionnaire was developed by Dodge and Coie (1987) as a teacher-report questionnaire. Parent-report and self-report questionnaires have since been developed (Kempes, Matthys, Maassen, Goozen, & van Engeland, 2006; Raine et al., 2006). These measures focus on identifying the occurrence of the behaviors within the past week or month, but offer little information about the psychological mechanisms involved.

Although reactive and proactive aggression predict different outcomes when measured with standard questionnaires, the two forms of aggression are highly correlated (see meta-analysis by Polman, de Castro, Koops, van Boxtel, & Merk, 2007, and other studies: Bailey, & Ostrov, 2008; Dodge et al., 1997; Hubbard, Parker, Ramsden & Smithmyer, 1998; Kempes et al., 2005). That is, individual children are more likely to have high levels of both reactive and proactive aggression than be high on one and low on the other. In fact, across studies, proactive aggression is rarely found alone, even in clinical groups, while reactive only or reactive plus proactive groups appear more consistently. Assessments that better distinguish between reactive and proactive aggression are needed in order to understand how each form of aggression leads to different psychosocial outcomes (Card & Little, 2006).

The SIP model has been used as way to better differentiate between the two types of aggression by distinguishing between the cognitive processes involved in the formulation of the aggressive behavior. The model has traditionally been measured through hypothetical vignettes. In a typical measure of the SIP model, children are asked

to imagine themselves in a hypothetical scenario in which the ambiguous actions of another child lead to a negative outcome (e.g.,, being pushed from behind which causes something cherished to be dropped and broken). Then children are asked a series of questions to determine their intent attribution (if the action was purposeful or accidental), their goals for the interaction (retaliation or preserving the relationship), their behavioral response (aggression or a socially competent response) and finally their evaluation of this responses (is aggression evaluated positively or negatively?). SIP vignettes have successfully been able to distinguish between reactive and proactive forms of aggression, allowing a lower level of analysis than questionnaires alone.

Both report questionnaires and hypothetical vignette measures of reactive and proactive aggression have been useful in characterizing these behaviors, but both measures have limitations. Retrospective report measures mainly focus on determining the past occurrences of aggression, with little focus on how these behaviors are generated in social interactions. Vignettes are a useful proxy for social interactions, but they do not capture actual behavior as it unfolds in real time and thus might not elicit the same kind of emotional and behavioral engagement as real situations do (Sharp, Ha, & Fonagy, 2011). How well a hypothetical scenario is able to bring online the cognitive mechanisms of the SIP model might also vary depending on how vivid one's imagination is and how much children would be able to "transpose" themselves into the situation described. In addition, vignette measures give the responder a prolonged time for considering their options, which would not be the case for most fast-paced real-world interactions (Yaros, Lochman, Rosenbaum, & Jimenez-Camargo, 2014).

Another limitation of these measures is that they fail to capture how aggressive behavior evolves over the course of an interaction, and how this evolution might be contingent on the responses of the interaction partner. Report measures are too broad to capture the fine details of how aggression waxes and wanes in an interaction. Even the steps described in the SIP model are only tested with regards to internal representations of the *initial* social stimulus (described in the vignette) that leads to the *first* behavioral reaction, but in real life, an interaction is unlikely to stop there. The vignette measures do not tell us what happens after the first aggressive response: what if it turns out the other child did not do it on purpose, or what if the other child reconsiders their actions and is trying to restore good relations – will the aggressor be appeased and return to formulating positive or socially competent behaviors? These kinds of dynamic responses are likely to vary for reactive and proactive aggression but few attempts have been made to elicit these responses in children.

Behavioral measures of aggression

In order to address these limitations, a few studies have used "real-time" tasks in which the play partner acts in ways meant to elicit frustration (e.g., Atkins, Osborne, Bennett, Hess, & Halperin, 2001; Atkins, Osborne, & Brown, 1993; Hubbard et al., 2002; Muñoz, Frick, Kimonis, & Aucoin, 2008; Phillips & Lochman, 2003; Waschbusch, Pelham, Jennings, Greiner, Tarter, & Moss, 2002). Four general experimental paradigms have been developed for aggression research. First, Hubbard et al. (2002) used a game in which a real play partner, a confederate child, cheats. They found that children high in

reactive aggression showed more angry non-verbal behaviors and higher skin conductance reactivity, with both increasing over the course of the game.

Second, a computerized pinball game with pre-programmed, anonymous partners has been used to distinguish between hostile and instrumental aggression. Players could frustrate the partner through either a hostile but non-instrumental action, playing an aversive noise in the partner's headphones, or an instrumental action that would block the opponent's game by "tilting" the partner's screen (Atkins et al., 1993; Atkins et al., 2001; Phillips, & Lochman, 2003). The measure was validated showing that the noise was perceived as hostile and the tilt move as both hostile and instrumental (Atkins et al., 1993). Atkins et al. (2001) showed that instrumental aggression in the game was correlated with teacher ratings of proactive aggression. Further, when the partner could punish aggressive behavior, continued aggressive responses were correlated with inattention and impulsivity. Phillips and Lochman (2003) used the same game in conjunction with an intervention designed to reduce aggression. Instrumental training (in which the child is rewarded for not being aggressive) reduced aggressive responses regardless of the children's reactive or proactive aggressive profile.

Third, a computerized competitive reaction time task has been used in several studies (Helseth et al., 2015; Muñoz et al., 2008; Waschbusch et al., 2002). Children competed with the partner to respond quickly to a cue on screen; the winner could take away points from the loser of each round and accompany it by a verbal message. Unknown to the child, the game was rigged and the opponent was automated: their decisions to take away points were predetermined and the messages prerecorded.

Children were thus exposed to several high provocation trials, in which the opponent took away lots of points and accompanied them with highly aversive messages, and some low provocation trials in which the opponent was more conciliatory. Waschbusch et al. (2002) found that children with a co-morbid diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) and ODD/CD reacted more to low provocation than the other groups (typically developing controls, ADHD-only and ODD/CD-only), while all groups reacted almost at ceiling to high provocation. The co-morbid diagnosis group showed a slower dissipation of anger and a propensity to hold a "grudge". In other words, they continued to punish the partner by taking points over multiple trials following high provocation. Muñoz et al. (2008) used the game with adolescents in a juvenile detention facility. They found that adolescents high in both reactive and proactive aggression were more likely to initiate provocation, whereas reactive-only adolescents did not. All adolescents responded to high provocation but the reactive-only group was highly responsive to low provocation as well.

Helseth et al. (2015) used the reaction time task with children with Conduct Problems (CP) and Callous-Unemotional traits (CU) adding another response option to the game inspired by Atkins et al. (2013), namely playing the aversive noise for the partner. They found that children with CPCU combined exhibited both proactive and reactive aggression to low provocation, while all children exhibited reactive aggression in response to high provocation. Also children with CPCU did not show the pattern of play indicative of holding a grudge found by Waschbusch et al. (2002) in children with co-morbid ADHD and CD/ODD diagnosis.

Finally, Yaros et al. (2014) developed a computer car race game for testing the SIP steps involved in evaluating potentially aggressive social interactions. The game included ambiguous provocations in the form of crashes from the opposing car. Children were asked to attribute positive (trying to be helpful) or negative (being hostile) intent to the opponent. They found that reactive aggression was positively related to hostile intent attributions when controlling for proactive aggression and proactive aggression was negatively related to hostile attribution when controlling for reactive aggression.

In summary, all of these tasks safely measure aggressive behaviors in an interactive context. However, the interaction contexts were purely competitive with no other incentives beyond winning the game. In these tasks the play partner always had the role of an opponent, and the games presented no opportunities to work *with* the partner. Also, beyond risking retaliation, acting aggressively in the game was not directly costly to the participants and there were no incentives to restore good relations with the partner. In real-life interactions, however, aggression can occur in non-competitive situations as well. Aggressors and victims often have more complex roles than being each other's adversaries. Further, in interactions beyond the game context, there are often costs to engaging in aggressive behavior: loss of resources, damaged relationships and missed opportunities for cooperating with others and pursuing mutual goals. Many of these limitations can be addressed by using "cooperative dilemmas" or games from behavioral economics and game theory. We next provide an overview of one such game, the Repeated Prisoner's Dilemma.

Background for the Prisoner's Dilemma

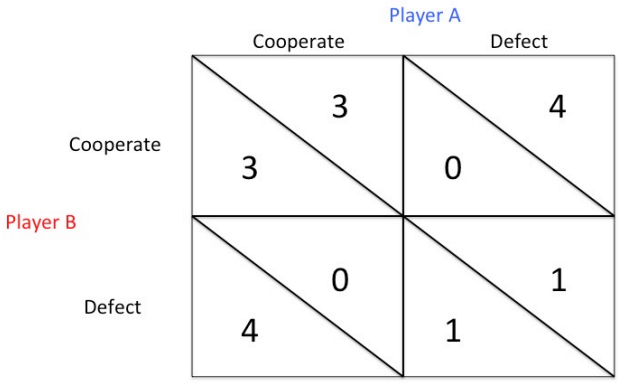
Cooperative dilemmas in the form of strategic games have been used to study live interactions for several decades in the fields of economics and game theory. These games provide an opportunity for studying both positive and negative interactions in real time and in a safe way. Games are typically played online, anonymity between players is maintained and real rewards can be gained or lost so that player decisions are consequential.

One key tool from game theory that has been used for measuring both positive and negative behavior is the Prisoner's Dilemma (PD) game (Rand & Nowak, 2013). As a task, the PD formulates a situation in which two people decide either to cooperate (C) or to defect (D), and the outcome (the payoff) is a result of their combined decisions (see Figure 1). The payoffs to the players are structured such that the combined payoff is maximized if both players cooperate (CC), but each player can potentially maximize their own individual payoff by defecting when the partner cooperates (DC). Defection can also protect one from receiving the lowest payoff: if one cooperates when the partner defects (CD) they get the lowest amount possible, often zero. Thus, both players have an incentive to defect (DD), but if this occurs they each get the same small amount.

The payoff amounts for the four possible outcomes typically falls in a rank order. For Player A (the first letter in the combination), the highest payout occurs when Player A defects while their partner cooperates (DC), a so-called exploitative move. The next highest is mutual cooperation (CC) followed by mutual defection (DD). The lowest payout, the so-called "suckers" payout, occurs if Player A cooperates and their partner

defects (CD). Both players typically face the exact same payout structure which is usually presented in a matrix (see Figure 1).

Figure 1. Example of payoff matrix for PD play



The PD has traditionally been used to study the conditions under which cooperation arises. The optimal outcome is for both players to cooperate (CC). Given the logic of the game, this move generates the most income overall and an equal or fair division of that payout. The evolutionary logic that underlies the game is that each person sacrifices the chance to earn the highest amount for themselves in order to create value (the highest total payout) and fairness. Dyads that are able to make this combined move tend to "survive" in mathematical models of evolutionary dynamics (Nowak, 2006). However, the cooperative move by each individual entails a risk – the partner could defect, leaving Player A with the suckers payout and decreasing their chances of survival (in the evolutionary games). Given this risk, when players in the PD interact only once, they tend to choose defection.

Several factors can improve the chances of cooperation in the PD, but the main one of interest here is repeated play with the same partner. In the Repeated Prisoner's Dilemma (RPD), partners play multiple rounds together, a scenario more similar to real-life, dyadic interactions. Repetition can lead to more cooperation but this depends on the probability of continuation, of playing another round with that partner. A longer number of rounds increases the probability that a player will cooperate, even in the first round of play. This so-called "shadow of the future" effect has been demonstrated in mathematical models (Axelrod & Dion, 1988; 1993; Fudenberg & Maskin, 1986) and also in live games with adults (Dal Bó, 2005; Dal Bó & Fréchette, 2011) and children (Blake et al., 2015).

Within an RPD, different forms of reciprocity, or "strategies" can arise which define how a particular individual plays the game. Some strategies are fixed – a player always cooperates or always defects – but most are conditional to some extent on the players' decisions in prior rounds. In the most commonly known strategy, Tit-for-Tat (TFT), the player cooperates on the first move and then matches their partner's last move on each subsequent round. Strategies can become increasingly complex in evolutionary models, but adults tend toward very basic strategies (Fudenberg, Rand & Dreber, 2012) and only a few will be examined in this thesis (see Chapter 3 and 4).

The use of strategic games to study psychopathology

Strategic games are starting to be recognized as useful measures for exploring social decisions and behaviors for clinical populations (King-Casas & Chiu, 2012)

including people with autism (Sally & Hill, 2006), anxiety and mood disorder (McClure et al., 2007; 2011; Rodebaugh et al., 2013), antisocial personality disorder (Rada, de Lucas Taracena, & Rodriguez, 2003), depression (Clark, Thorne, Hardy & Cropsey, 2013), psychopathy (Morkos et al., 2008, Rilling et al., 2007) and conduct problems (Sharp et al., 2011). However, only a few of these studies use repeated games or multi-move games and most are with adults.

To date, only a handful of studies have used the RPD and similar interactive economic games (i.e.,, Trust Game) to study psychopathology in adolescents and children (Sharp, 2012), yet the results have been promising. For example, anxious adolescents playing an RPD were found to be more cooperative than healthy controls both after partner cooperation (McClure et al., 2007) and partner defection (McClure-Tone et al., 2011), suggesting a greater desire to establish and maintain positive interactions. By contrast, adolescents from a community sample with higher externalizing problems were more likely to defect and take advantage of a partner's cooperation (Sharp et al., 2011). One study testing several economic games, including the RPD, in children with autism spectrum disorder found some differences in strategies used by this group compared to typically developing controls (Sally & Hill, 2006), but the overall sample was small. More recently, using a novel, child-friendly interface for the RPD, Blake et al. (2015) found that middle-schoolers who scored in the abnormal range of the Conduct Problems sub-scale of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) cooperated less after their partner's defection. Children with no conduct problems on the

other hand were more likely to re-established cooperation both after the partner's defection and after they had exploited the partner (DC) in the last round.

Combined, these results suggest that RPD can elicit safe negative interactions driven by real incentives and can be used as a live, child-friendly measure of social interactions. Because children engage in contingent behavior in the RPD game (Blake et al., 2015), the task provides an opportunity to investigate how negative behaviors evolve over the course of the interaction. These modulations can be detected through different strategies and patterns of play throughout the game. Thus, by linking play strategies in the RPD game with reactive and proactive behavior outside the game, much can be learned about how children showing different types of aggression modulate their behavior in social interactions with different kinds of partners.

The RPD thus addresses several of the limitations described for other tools used to study social interactions: 1) it allows for direct measure of behavior in a real and incentivized interaction; 2) unlike vignette-based measures which only describe one action and allow one response, the RPD allows for multiple inputs from the partner and multiple responses over the course of the game; 3) unlike other games the RPD allows for varied inputs (both negative and positive) from both the partner and the participant and thus makes investigation of negative interactions possible beyond the immediate response following provocation; 4) finally, as opposed to other competitive tasks (e.g., Waschbusch et al., 2002) in which the player only gets feedback from the partner when they lose, RPD is a more dynamic and realistic interaction where the signal from the partner is ongoing.

However, existing versions of the RPD for children have limitations. Most have used live play between children mirroring games done with adults, which includes random pairing of players. This limits control over the experience that each child has. To address this limitation, some tasks have used standard strategies, for example Sally and Hill (2006) used the TFT strategy played by a confederate. But this too limits ability to test how different children respond to similar behaviors by a partner. For example, a TFT partner starts by cooperating and then mirrors the child's decisions, so any defection in the game must be initiated by the child. This means that children who also play TFT will never experience partner defection whereas a child who defects may end up in a cycle of defection with the partner. To properly address this limitation the development of a novel version of RPD is needed in which fixed patterns of play are designed to create key circumstances for testing responses. There is also need for testing this novel approach to ensure that children engage in strategic play when the play-partner does not act contingently (follows a fixed sequence of decisions regardless of what the child's moves are). Another limitation of current RPD tasks is that little is known about how children construe the interaction and what kind of cognitions lead to their decisions in the game. When asked about their PD interactions, adults provide responses that describe strategic play (Fudenberg et al., 2012), but to our knowledge there is no such previous data for kids. A key goal of this dissertation was the development and validation of a new RPD that addresses these limitations.

The Present Research

This dissertation integrates approaches from developmental psychology and behavioral economics and game theory to accomplish the following goals: 1) to understand how children interpret the repeated RPD interaction and what meanings they attribute to their own and their partner's decisions in the game (Chapter 3); 2) to use the RPD task to investigate the reactive and proactive aggressive behavioral phenotypes by predicting different patterns of play in the RPD for each form of aggression (Chapter 4); and 3) to expand the SIP model to novel positive interaction vignettes and assess whether a) negative and positive responses are generated in ways expected by SIP and whether b) the SIP results can explain positive and negative behavior in the RPD (Chapter 5).

Chapter 3 shows evidence that the RPD task can be used as a minimal, quantifiable social interaction with children. The task elicits rich social cognitions, and can be used to measure patterns of social behavior through strategies of play. Classes of motivations such as interpersonal and strategic self-interest predicted RPD behavior. Chapter 4 shows that RPD measurements can be applied to aggressive behavior, and that decisions in the RPD task can reveal new aspects of the reactive aggression in social interactions: that high-reactive children followed a game-theoretic strategy known as GRIM – they did not return to cooperation after partner defection. Finally, Chapter 5 describes how responses to positive interaction scenarios were consistent with the SIP model but did not predict RPD play, offering partial evidence to the hypothesis that cognitive mechanisms that have been used to explain the occurrence of aggressive behavior can be extended to explain positive social behaviors like reciprocity as well.

CHAPTER 2

PARTICIPANTS, METHODS AND MEASURES: DEVELOPING A FIXED STRATEGY RPD TASK FOR CHILDREN AND EXTENDING THE SIP MEASURE FOR POSITIVE INTERACTIONS

To study both positive and negative interactions by combining approaches from developmental psychology and behavioral economics, we redesigned the classical RPD and SIP measures. The novel versions add key features, such as partners with fixed strategies for the RPD, and positive-outcome vignettes for the SIP measure. Fixed strategy partners allow for a standardized experience in the game as well as tests of responses to particular patterns of play; the positive-outcome SIP vignettes enable the testing of the SIP model for positive social behaviors, and possible connections with the RPD measure could validate the model for cooperative social interactions. This chapter describes the measures, the study procedure and the study participants.

Participants

Children between the ages of 9 and 11 (N = 167, Females = 82) participated in this study either online, from home (N = 109), or online at school (N = 58), in their classes. Children were tested in schools, in the presence of an experimenter, to check for differences between home and school online play and to confirm that children participating from home were engaging with the task in the same way. Students at two US elementary schools participated in the study: a technology charter school serving a

generally low SES community and mostly minority children ($N = 22$), and a private school for gifted children ($N = 36$) from upper middle class highly educated families. Both schools focused on children gaining experience in using computers as early as the first grade, and both schools were able to provide laptops for students to access the online testing platform. Children who participated online from home were recruited through a database of families who had previously participated in research in our university's Child Development Labs. These families tend to be upper middle class and highly educated similar to those served by the private school mentioned above. Parent consent was obtained prior to testing, either through the online testing platform (for children participating from home) or through a written consent form sent home to the parents (for children who participated from school). Child assent was obtained through the online testing platform. All consent, measures and experimental procedures were approved by the Institutional Review Board at Boston University (IRB #3934E).

The overall mean age was 9.96 years ($SD = 0.83$). We selected this age range because previous studies suggest that at ages 9 through 11 children are able to play the PD game in a strategic way: making intentional contingent decisions and being able to modulate these decisions based on those of the partner (Blake et al., 2015; Matsumoto et al., 1986; Perner, 1979). Also, for most children, this period marks the end of the transition from physical manifestations of aggression to social manifestations of aggression (Vitaro et al., 2006), the latter being the focus of this study. An additional 7 children did not finish the study or data was not recorded due to technical errors.

Procedure

For those children who participated from home, all the measures were administered online. Following informed consent parents reported on their child's aggressive behaviors by answering the Parent Rating Scale for Reactive and Proactive Aggression (PRPA; Kempes et al., 2006). Parents were then asked via text instructions to help get their child set up at the computer. For children who participated in school, consent forms were sent home to the parents and later retrieved with the help of the teachers. On the testing day, an experimenter helped children to get set up at the computer and access the online data collection platform, and remained in the classroom for assistance for the duration of the testing. All children had to give their assent to participate on their computers to be able to access the online measures.

Children first completed the SIP measure (described in detail below) through the online platform, including vignettes with both negative and positive outcomes. After reading each vignette children answered questions related to the SIP steps.

Following the SIP measure, children were introduced to the RPD game. In order to incentivize them to take the game seriously, children were shown a variety of e-gift prizes (games, e-books, etc.) and were told that they could use the points earned in the game to choose and redeem some prizes. After practice trials and comprehension checks for the RPD (see details in Appendix A), children played three RPD games against different partners. Children were told that they were being connected online to other participants. Children who participated in schools were led to believe that they would be playing against children accessing the study from their homes. Unbeknownst to the

children, partners were pre-programmed algorithms using a particular strategy during the game: a TFT partner, a mostly cooperative partner, and a mostly non-cooperative partner. After playing with all three partners, children answered questions about how they and their partner played the game. Based on the number of points earned in the game, children selected prizes at the end, which were sent to their parents' e-mails for redeeming. After the experiment ended, parents were sent a debriefing form describing the deception in the game (the partner was not a real child) and why it was necessary. In the schools, the debriefing was done by the experimenter who then lead a discussion about why the deception was needed.

Measures

The SIP questionnaire

The standard SIP questionnaires consist of vignettes that describe negative interactions and outcomes followed by questions about the different steps in processing this social information. We focused on three of the five SIP steps: interpreting the situation by attributing intent, generating possible behavioral responses and evaluating the responses, since these are the steps that have primarily been used to distinguish between different types of aggression, reactive and proactive, and because these steps were feasible to test with a multiple-choice questionnaire, delivered in written format. We designed a similar questionnaire drawing examples from various other SIP assessments (Bell, Luebbe, Swenson, & Allwoold, 2009; Camodeca & Goossens, 2005; Flood, Hare,

& Wallis, 2011; van Rest, van Bokhoven, van Nieuwenhuijzen, Embregts, Vriens, & Matthys, 2014; see Appendix C).

Three negative vignettes were used, depicting three interactions in which another child produced a negative outcome for the participant (muddy homework, dropped ice-cream and tripping). In the tradition of the SIP questionnaires, the action was presented in an ambiguous way, and the child was asked to interpret whether the outcome was accidental or intentional by indicating whether the child in the story did it on purpose and whether the child was trying to be mean. Then the participant was asked how likely it was for them to respond aggressively in that situation by saying or doing something mean to the perpetrator. We asked the participants both about an immediate response ("right away") and a distant one ("in the future"). Finally, inspired by measures designed by Camodeca and Goossens (2005), and Kupersmidt et al. (2011) we asked participants to evaluate a potential aggressive response (both an immediate and a distant one) in terms of how easy it would be for them to respond aggressively (self-efficacy), the consequences of the response (how well things would turn out if they responded aggressively) and how morally acceptable an aggressive response would be (would it be good or bad to act aggressively).

In addition to these negative stories, we also created three novel positive stories, which allow the same questions to be asked. Similarly, these stories presented an ambiguous interaction in which another child produced a positive outcome for the participant (a wanted object is brought close to the participant, a ball is kicked in the participant's direction as they are contemplating joining the game, and a seat is made

available when the participant needs it). Participants were asked whether the actions in the stories were accidental or intentional. Then they were asked about the likelihood of responding positively (right away or in the future) by doing or saying something nice to the other kid involved in the interaction. Finally, mirroring the negative vignettes, participants were asked to evaluate a potential positive response in terms of the ease of carrying it out, the consequences and the moral value of such a response.

All responses were given on a four point Likert scale and scores for intent attribution, response generation and the three dimensions of response evaluation (self-efficacy/ease, consequences, and moral acceptability/value) were calculated by averaging the responses to the questions corresponding to them separately for the negative and positive vignettes. For example, answers for the questions: "How much do you think the kid did this on purpose?" and "How much do you think the kid was trying to be mean?", which test for hostile attribution bias, were averaged together to create the hostile attribution score (for a list of all questions grouped under corresponding steps and the multiple choice answers see Appendix D). The minimum score possible was 0 and the maximum 3.

We were unable to calculate positive intent attribution scores for 2 children and positive response scores for 2 children due to missing data. Missing data for 2 children also prevented the calculation of the response evaluation score for negative vignettes and missing data from 4 different children prevented the calculation of the response evaluation score for positive vignettes.

Reactive and Proactive Aggression Measure

Aggression was measured through parent report. Parents responded to the 11-item Parent Rating Scale for Reactive and Proactive Aggression (PRPA) developed by Kempes et al. (2006) rating the frequency of behaviors described in the items for their child (never = 0; sometimes = 1, often = 2). Reactive aggression and proactive aggression scores were calculated according to questionnaire instructions. Given that we tested a community sample, the children were grouped into low and high aggression profiles based on a median split (see details in Appendix C). Due to missing data we were unable to compute the reactive aggression score for 14 children and the proactive aggression scores for 9 children. Out of these children 8 were missing scores for both proactive and reactive aggression. Thus for analyses involving only the reactive aggressive score, data from 153 children was used. For analyses involving only the proactive aggressive score data from 158 children was used. For analyses involving both reactive and proactive aggressive scores data from 152 children was used.

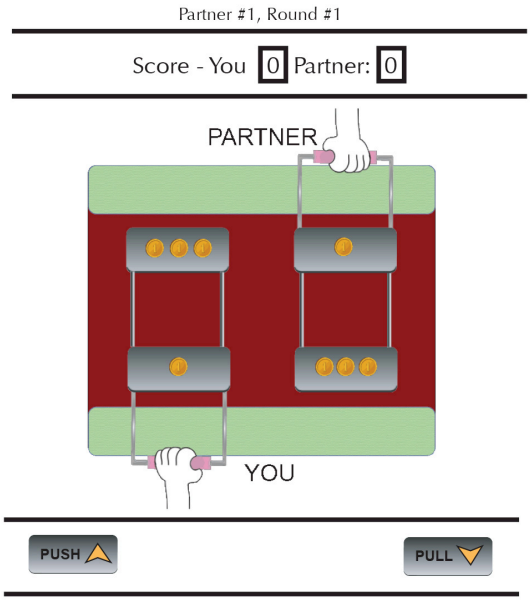
Similar to previous studies (Kempes et al., 2006; Muñoz et al. 2008) we grouped children into different aggression groups. On the reactive aggression scale (6 items), scores ranged from 0 to 10 (max score of 12) with a median score of 3. Children below or equal to the median score ($N = 95$) were categorized as low reactive aggressive, while children whose score was above the median ($N = 58$) were categorized as high reactive aggressive. On the proactive aggression scale (5 items), scores ranged from 0 to 6 (max score of 10) with a median score of 1. Children who scored below or equal to the median versus above the median were categorized as low ($N = 125$) and high proactive

aggressive ($N = 33$), respectively. The relatively low size of the high proactive aggression group is consistent for studies using community samples (e.g., Bailey & Ostrov, 2008). Note that the two categories have unequal number of children because children obtaining a score equal to the median (reactive aggression: $N = 19$; proactive aggression: $N = 47$) were categorized as low reactive and proactive aggressive respectively.

Repeated Prisoner's Dilemma (RPD) game

The RPD game was played using a novel online game platform (see Figure 2), modeled on a visual interface for the RPD that was used successfully with 10-11 year olds (Blake et al., 2015). The new version of the game used a similar interface but was created for easy online access, and designed as a pre-programmed partner game as opposed to a live, child-to-child game. The game was developed, pilot-tested in the lab with undergrads and improved over several months.

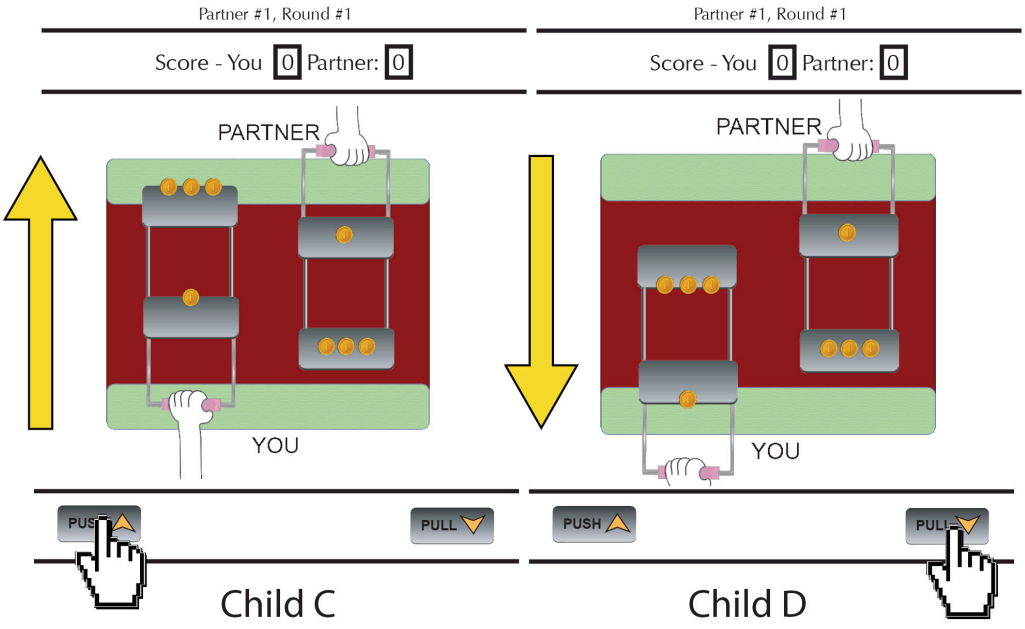
Figure 2. Initial display of the RPD game board



In the RPD game task, children faced a situation in which they had to either cooperate (C) with their play partner or defect (D). To avoid priming the children, the task was never described as a cooperation task and the actions in the game were never referred to as cooperating or defecting. Instead, the visual interface presented two buttons labeled Push and Pull, representing the cooperative and defecting moves, respectively.

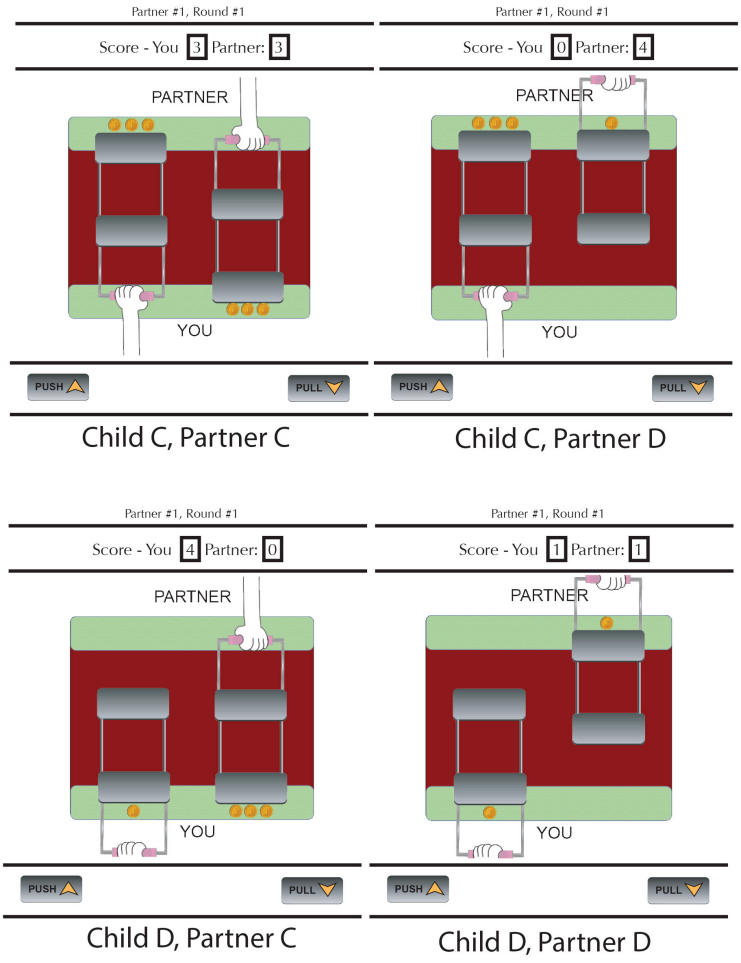
If the child clicked Push, 3 coins were delivered to the partner and the one coin closer to them fell into the "abyss," a red zone in the middle into which coins fell and then disappeared. Clicking Pull resulted in the 3 coins on the far side falling into the abyss and the one coin delivered to the child. This set-up mirrored the classical payoff matrix for the Prisoner's Dilemma.

Figure 3. Action options in the RPD game



If both children cooperated, they each received 3 points. However, if one player defected and the other cooperated, the defector got 4 points and the cooperator got 0 points. Thus, the child had an incentive to defect, but if both partners defected, they each got 1 point. Decisions in the game were made simultaneously by the two players (child and programmed partner) and the consequences of all decisions were made visually obvious and intuitive. The numeric outcomes of each round were shown on screen and added to a running point total at the top of the screen.

Figure 4. Outcomes in the RPD game



Children played a total of three RPD games, each against a new, anonymous partner for ten rounds with each partner. The partners in the game were pre-programmed strategies, partly based on prior research with adults and designed to elicit different responses from the children. However, children were told that each partner was another child. The first partner used a Tit-for-Tat (TFT) strategy which cooperates in the first round and then mirrors the child's decision in the prior round: cooperating if the child cooperated on the previous round, or defecting if the child defected on the previous round. The next two partners were counterbalanced across participants. One partner cooperated 80% (80C) of the time, defecting on rounds 3 and 7. A third partner defected 80% (80D) of the time, cooperating on rounds 3 and 7. There were ten rounds for each partner to allow for a mix of C and D trials in the case of the fixed strategy partners. The final round was announced in each game in order to give children the opportunity to defect in the final round, a typical strategic move from game theory.

The order of the three partners was partially counterbalanced. The TFT partner was always first, followed by either 80D or 80C. Playing against a TFT partner was meant to capture the child's baseline play strategies, since this partner was programmed to simply "react" to one's decisions on the previous round. The 80D and 80C partners were designed to "provoke" through their play strategies and elicit both cooperation or defection.

Motivations to cooperate or defect in the RPD

After playing all of the RPD games, children answered three open-ended questions about their own decisions and those of their partner. The questions were designed to elicit their spontaneous thoughts about the game, focusing on what motivated cooperation (push) and defection (pull) decisions. One question focused on the child's own play:

1. We are interested in how you played the game. Please describe why you clicked the push and the pull buttons.

This question asks about both cooperation and defection since it is possible that some children only use one of the two options. Two separate questions were asked about children's perceptions of their partner's motivations to defect and cooperate:

2. Sometimes your partner pressed "pull". Why do you think they did that?

3. Sometimes your partner pressed "push". Why do you think they did that?

These three questions appeared in a fixed order for all children.

Stated goals for the RPD game

Children next answered a multiple choice question which was intended to capture their goals for the game, as well as secondary motivations that they might have not offered spontaneously:

People think about different things when they play this game. We are interested in what you were thinking about when playing. Please check all of the boxes that describe your thoughts during the game:

a) I wanted to win as many total points as possible

b) I wanted to win points every round

c) I wanted to win more points than the other kid

d) I wanted to help the other kid win points

e) I wanted to punish or reward the other kid for their choices

f) I had no particular reason in mind. I just pushed the buttons randomly

Since the RPD game is designed to pit different goals against each other, children were allowed to select more than one answer. The answers were inspired by work with adults (Dreber, Fudenberg, & Rand, 2014), and were meant to capture the following goals: maximizing one's own gain, loss avoidance, competition, prosociality, and reciprocity.

The following chapters show how these measures were combined and used to investigate different aspects of social interactions.

CHAPTER 3

MOTIVATIONS, DECISIONS AND STRATEGIES IN THE REPEATED PRISONER'S DILEMMA GAME

Although extensively researched in adult experiments, the RPD has only been used with children in a small number of studies (e.g., Blake et al., 2015; Matsumoto, Haan, Yabrove, Theodorou, & Cooke Carney, 1986; Sally & Hill, 2006). The game establishes a minimal social interaction suitable for studying social behaviors, but little is known about how children construe this interaction and what kind of cognitions lead to their decisions in the game. Also, the dynamics of the game and strategies of play have only recently begun to be explored with children (Blake et al., 2015). This chapter presents findings on how children's self-described motivations in the game relate to their cooperative behavior, and how children's decision patterns resemble standard strategies from game theory.

Although many studies have tried to infer motivations in strategic games using clever experimental manipulations (see review by Kopelman, Weber, & Messick, 2002), very few have explored directly how people think of these games by asking them to explain their decisions. Examining people's stated motivations can offer a wealth of information about how the task is construed, and whether it elicits the types of social cognitions that researchers are trying to investigate. Dreber et al. (2014) have studied when and why adults cooperate in the RPD in this direct way. They asked participants to indicate how likely they would be to cooperate based on the outcomes in the previous

round, and their motivations for cooperating. They found that, for adults, earning the most points in the long run was the most important motivation for cooperating in the game.

To our knowledge, no studies have simply asked children how they play the RPD and about their motivations for different decisions. One study by Sharp et al. (2011) did ask children about their decisions in a related game, the Trust Game. This game involves two players, one of whom (the investor) makes a first move, deciding whether to keep money or pass it on to a trustee. The money sent would get tripled and the trustee would then decide how much money to send back to the investor. Boys with and without externalizing behavior problems played the trust game and were asked to explain what motivated their decisions in the game. Responses, as well as actual decisions, revealed that children with externalizing problems were both less trusting and less trustworthy.

The RPD game originated in the context of game theory and has been used to study mathematical models of how cooperation emerges between rational decision-makers. A wide variety of strategies have been extensively tested in tournaments where computer algorithms (defined by the different strategies) play against each other. In the most famous of these tournaments (Axelrod & Hamilton, 1981), TFT emerged as the winner against far more complicated strategies. Research with adults shows that decision patterns in the RPD sometimes approximate those of the standard strategies from game theory (Dal Bó & Fréchette, 2011; Dreber et al., 2014). By studying how and when people's patterns of play align with these standard strategies, one can better understand

the social cognitions that drive their behavior, such as forgiveness, leniency and exploitation (Dreber et al., 2014).

Approximation of standard game theoretic strategies has never been rigorously calculated for children's play of the RPD, and it is not even clear that children consciously use particular strategies. One prior RPD study found suggestive evidence that girls and children with few conduct problems used a strategy similar to the Win-Stay-Lose-Shift (WSLS) strategy (Blake et al., 2015). In WSLS, players repeat their prior move (stay) if they earn either of the top two payouts (CC or DC), but change to the opposite move (shift) if they receive the bottom two payouts (DD or CD). In contrast to this result, boys and children with conduct problems appeared to follow a GRIM strategy. For GRIM, players cooperate until the partner defects and then they defect continuously, never attempting cooperation with that partner. While this strategy appeared plausible, the high use of ALL-D (always defect) may have distorted the results given the small sample. Further, no formal calculations were done to determine how closely these strategies were followed.

Given the limited data on how children understand strategic games, we included questions after children had played all of the RPD games. Specifically, we asked how they approached the task, what their own motivations were, and what they believed about the partner's motivations. These questions were intended to provide a critical validation that a) children view the RPD as a real social interaction, b) their explicit thoughts about the game would predict actual decisions in the game c) they actually think strategically about their decisions. As validation of strategic thinking, we analyzed how well

children's patterns of play aligned with three standard game-theoretic strategies in the RPD: TFT, WSLS and GRIM.

Results

Children's motivations for cooperating or defecting in the RPD

Children expressed a variety of motivations for both cooperating and defecting and identified similar motivations for the decisions of their partners in the game. Many of the motivations expressed by children matched well-studied factors that have been shown to modulate responses in PD games with adults (Kopelman et al., 2002). The RPD game, by design creates a situation in which motivations such as self-interest, get pitted against the need for establishing interpersonal cooperation by working together with the partner. Sometimes, self-interest can even give way to prosocial motivations, such as the desire to help the other player win points or achieve an equal outcome. The open-ended answers to our questions allowed us to capture diverse and rich nuances for all these different motivations.

We coded the children's answers for three main types of motivations: strategic self-interest, interpersonal or prosocial as well as motivation subtypes (see Table 1 and details about the coding scheme in Appendix B). Answers were assigned a code of 1 for each type or subtype of motivation the child mentioned and a code of 0 for the rest. Children who mentioned more that one motivation received a code of 1 for each type or subtype they mentioned. The answers were coded by two independent coders who were

blind to any other information about the child including, their age, gender and decisions in the game. Average Cohen's kappa for inter-coder reliability was 0.72.

Strategic self-interest

Motivations indicating strategic self-interest appeared most often in children's explanations of their decisions in the game (mentioned by 45.5% of children), and particularly in their explanations of the partner's defection in the game (78.4%). A majority of children attributed strategic self-interest motivations to partner's defection.

Table 1. Children's motivations for decisions in the game

	Child motivations	Partner motivations
Strategic self-interest	**45.51%**	**81.44%**
Personal gain	40.71%	79.64%
Freeriding	6.58%	6.58%
Interpersonal	**35.33%**	**37.12%**
Trust	3.59%	12.57%
Reciprocity	22.75%	19.76%
Fairness	11.37%	8.98%
Joint gain	3.59%	9.58%
Prosocial	**10.77%**	**50.30%**
Random	2.99%	5.99%
Other	16.76%	19.76%

* Percentages will not add up to 100% because some children mentioned multiple motivations

Two subtypes of strategic self-interest were identified: the desire to maximize personal gain and the intention to freeride, take advantage of the partner. While both have self-interest as the end goal, the latter focuses more on the interaction in the game.

Personal gain. These answers focused on increasing personal gain through acquiring points, winning, or getting prizes. Eight responses referred to relative gain by preventing the partner from getting more points (e.g., "I clicked only the pull buttons because my partner would get more points if I clicked the push button") and were also included in this category. Most children viewed defection as a way to increase personal gain, and only one child mentioned cooperation (push) as a sometimes better alternative to pulling (defection) for increasing one's score: "I clicked pull because I'm always safe, either I get one point or four points. I clicked push because I could also get there, and that's better than one."

Although all answers in this category focused on personal gain as a motivation, they differed in the degree to which personal advantage was at play. Some children used defection to get points when they were losing, as a way of catching up (e.g., "when I'm far behind I pull to give myself a boost"), others were extremely competitive and gain-oriented (e.g., "I did pull a lot because GO BIG or GO HOME! And I never EVER lose!"). Kids also differed in their reasons for defecting on their partners, with some choosing it for safety (e.g., "I clicked pull so I got guaranteed points") others as a way to get more points than the partner (e.g., "I wanted to get as many points as possible and get

more points than my partner"). Similarly, in attributing motivations to their partners, children identified personal gain as the main motivator for defection, some believing the partner defected "to get more points" while others felt the partner was defecting "so that they didn't just get nothing, even if it was only one point". Some children even attributed negative personality traits to their partners, such as greed, selfishness or being a cheater (e.g., "they were trying to be selfish (but I'm not being accusatory)"). Some children felt that partner's defection was purposefully obstructive and meant to prevent them from winning (e.g., "to keep me from earning points" or "to be annoying").

Freeriding motivation. Some children focused on tricking or taking advantage of the partner as a motivation for decisions in the game (e.g., "I clicked the push button to lure my opponent into giving points to me and then I kept pulling.") A small number of children ($N = 2$) even described defection in the last round, a well-studied move in game theory, as part of their strategy: "If they pushed, we would push until round ten, when I pulled because they could not do anything to take from me in a future round because it was the final." Children also sometimes felt deceived or taken advantage of by their partners, with some children interpreting partner defection as an attempt at freeriding (e.g., "because I press push a lot so they wanted to get 4 points").

Interpersonal motivations

Many children gave interpersonal motivations for their decisions (35.33%), both for cooperation and for defection, explaining their own decisions in relation to those of their partner. Children attributed interpersonal motivations to their partners as well

(37.12%), mostly when explaining partners' cooperation. Interpersonal motivations were subdivided into four topics: trust, reciprocity, concern for fairness and joint gain.

Trust. For some children trust was a reason for cooperating or defecting. Children mentioned trusting the partner or not (e.g., "we trusted each other we both clicked push every time"), wanting to make the partner trust them (e.g., "I usually kept pressing push to get my partner to trust me") or they expressed the desire to use a decision to test the partner's trustworthiness (e.g., "First, I always pressed push so I could see if my partner liked to share or didn't"). Some children also felt that trust played a role in their partners' decisions, with some partners cooperating "to help build trust", others trying to gauge how trustworthy they were (e.g., "because he was testing me"), and some defecting because of lack of trust (e.g., "maybe they did not trust me to click push").

Reciprocity. Children mentioned engaging in both negative (e.g., "whenever someone pushed pull I would pull the next round to even the playing field") and positive reciprocity (e.g., "if the person gave me 3 coins I give him/her back 3 coins"). Children mentioned both responding in kind to the partner (e.g., "I clicked pull to give me points when I saw the other person was pulling") and trying to elicit a response in kind from the partner (e.g., "I clicked push so that my partner would give me some coins"). The answers children gave suggest that the situations in which they chose to reciprocate varied. Some children seemed to describe a pure TFT strategy (e.g., "I clicked whatever my partner clicked the round before"), other children resorted to responding in kind only after the partner's repeated defections (e.g., "if they kept on doing pull I would switch to pull for a while"). Some answers suggested that reciprocity was a deviation from the

default decision (e.g., “I clicked push most of the time until my last partner clicked pull, pull, pull, pull and pushed once or twice it made me VERY MAD”, or “I only pushed if my partner did too”) and that they did not engage in perfect reciprocity across the game (e.g., “I wanted to get as many points as I could. If someone did something nice for me, I USUALLY paid them back”).

Children felt that their partners also engaged in reciprocity. Children attributed positive reciprocity to their partners (17.36%) mentioning that the partner had cooperated in response to their own cooperation (e.g., “I think my partner pressed push because I pressed push a lot.”), or to try to elicit cooperation (e.g., “so that I would be nice in return and give them coins”). Only a small percentage (2.39%) of children felt that partner’s defection was in response to their own (e.g., “because I pulled so they wanted to get points so the next round they pulled because they thought I was being greedy”).

Fairness. Some children’s answers suggested a preoccupation with playing a fair game. These answers focused on sharing and reaching an equal (or closer to equal) outcome as motivators for their decisions. Some children expressed a desire to have an equal amount of points with their partner (e.g., “I clicked both push and pull buttons and tried to make it so scores were the same and fair”) or for their scores to be more similar (e.g., “I would push sometimes when I was winning by a lot because I felt bad for them and so it would be a closer game”). Children viewed both cooperation and defection as a means to playing a fair game. Some children even mentioned taking turns defecting as a way to distribute points fairly (e.g., “I clicked pull when my partner clicked pull and we took turns giving each other coins”). Children also felt that their partners were concerned

with fairness, especially as a motivation for cooperation (e.g., “I think they agreed with my idea of both of us pressing push so it would be more fair”).

Joint gain. A small percentage of children (3.59%) mentioned their decisions being motivated by joint benefit from collaboration. Few spontaneously mentioned that by working with their partner one could increase the amount of points jointly earned (e.g., “I wanted to get the most points but I needed my partner to help me. So I pushed my 3 to him and he pushed his 3 to me”). More children (9.58%) attributed joint gain as a motivator for partner cooperation (e.g., “So that we each get three points, everybody wins”).

Prosocial motivations

Prosocial motivations expressed by children were distinct from interpersonal ones because although the focus remained on the partner, they were not conditional on the partner’s decisions. Children who invoked prosocial motivations for cooperating mentioned empathy, generosity and the desire to be nice or to help the partner (e.g., “I wanted to be nice so I mostly clicked push”). In all of these instances cooperation was not intended to trick the partner or to determine the partner to cooperate. Answers indicate that the frequency of prosocial cooperation over rounds varied. Some children suggest that they cooperated most of the times, others that they were only sometimes compelled to cooperate as a gesture of kindness (e.g., “sometimes I felt nice and empathetic”). A large percentage of children (50.30%) interpreted their partners’ cooperation as having prosocial motivations, making this the main perceived motivation for partner cooperation.

Some children even suggested that the partner intended for them to win (e.g., "to help me win").

Finally, in addition to children mentioning strategic self-interest, interpersonal and prosocial motivations, there were also children who said that they made their decisions randomly (e.g., "I wanted to see what happens when you click either button") and others gave nonsensical or non-informative answers like describing the rules of the game or what their decisions were but without explaining why.

Motivations and decisions in the PD game

To examine the relation between children's motivations for their own decisions and their actual decisions in the PD game, we used a nested logistic regression model (Stata v.14.2). The PD game produces a binary decision in each round: Cooperate (coded as 1) or Defect (coded as 0), thus the use of logistic models. Each child played thirty rounds in total, which were treated as a repeated measure with rounds nested under each participant. The slopes and intercepts were allowed to vary for each participant with clustering of standard errors. The three main types of motivations: strategic self-interest, interpersonal and prosocial (coded as 1 if present in the child's answers, 0 if absent) were introduced as independent variables in the model. Age (in years) and gender were introduced as covariates.

On average children cooperated 35.7% of the times across partners. Children who mentioned interpersonal motivations for their decisions were significantly more likely to cooperate ($B = 0.951$, $p < 0.001$) than those who did not (see Figure 5 and Table 2). The

likelihood of cooperation was significantly reduced for children who mentioned strategic self interest as a motivation for their decisions (B = -0.568, p = 0.001). Prosocial motivations however were not significantly related to cooperation (B = 0.219, p = 0.323).

Figure 5. Motivations and overall cooperation in the RPD. The child's stated motivations for their own decisions are on the left and their beliefs about the partner's motivations are on the right. Average cooperation is included as a line for reference.

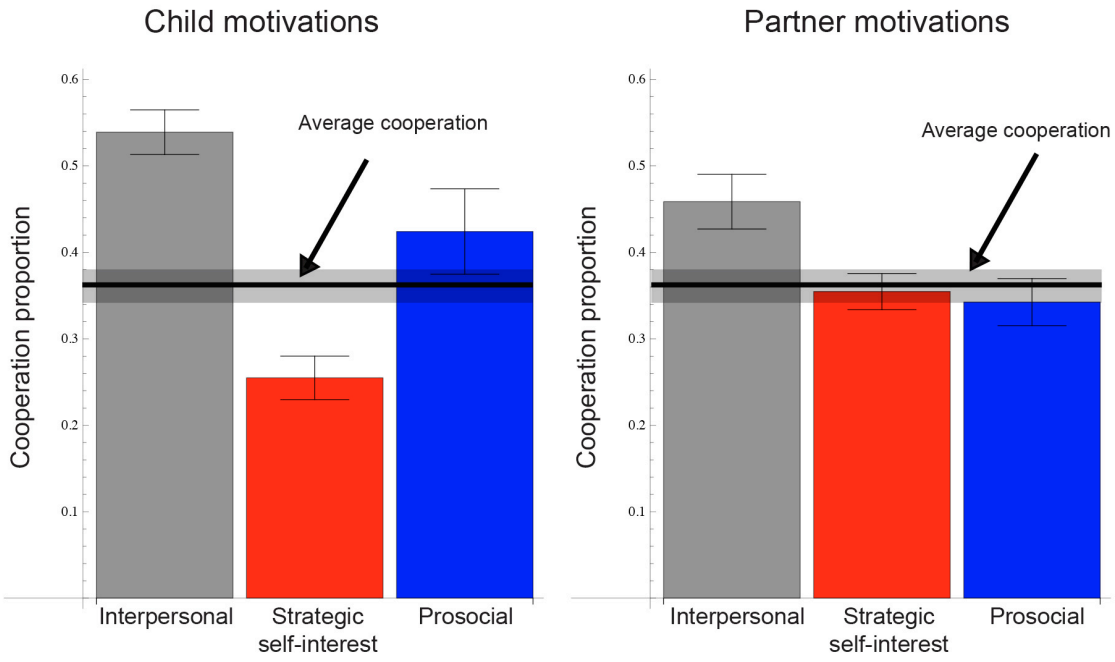


Table 2. Relations between children's motivations and their decisions in the RPD

	B	*SE*	OR	95% CI	
Intercept	-2.867	0.959**			
Age	0.216	0.096*	1.24	1.03	1.50
Gender	-0.052	0.152	0.95	0.70	1.28
Interpersonal motivations	0.951	0.170***	2.59	1.85	3.62
Strategic self-interest	-0.568	0.172**	0.57	0.40	0.79
Prosocial motivations	0.219	0.222	1.24	0.81	1.92
Log Likelihood	-3019.142				
# Observations	5010				
# Children/Clusters	167				
χ2(df)	96.82(5)				
Pseudo R2	0.08				

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We also tested whether the more common motivation subtypes predicted the likelihood of cooperation. To avoid collinearity (remember that children sometimes mentioned multiple motivations) we used separate logistic models to test these predictors. Reciprocity, the most commonly mentioned interpersonal motivation significantly predicted an increased likelihood of cooperation ($B = 0.973$, $p < 0.001$). This was true for both positive reciprocity (motivation to cooperate in response to partner cooperation; $B = 0.493$, $p < 0.016$) and negative reciprocity (motivation to defect in response to partner defection; $B = 1.266$, $p < 0.001$). Fairness, the second most commonly mentioned interpersonal motivation was also significantly related to an increase in the likelihood of

cooperation ($B = 0.473$, $p = 0.022$). Personal gain significantly predicted decreased cooperation ($B = -0.835$, $p < 0.001$).

Age was also a significant predictor of cooperation with older children being more likely to cooperate ($B = 0.215$, $p = 0.025$). We also investigated whether age and gender were related to the types of motivations that the children mentioned for their own decisions. Older children mentioned interpersonal motivations more often ($B = 0.660$, $p = 0.001$) as well as prosocial motivations ($B = 0.659$, $p = 0.037$) while there were no age differences in children mentioning strategic self-interest ($B = -0.116$, $p = 0.539$).

To examine how children's decisions were impacted by the kinds of motivations they attributed to their partner we conducted a nested logistic regression with the partner motivations (strategic self-interest, interpersonal and prosocial) as independent variables (see Figure 1 and Table 3). Attributing interpersonal motivations to the partner increased the likelihood of cooperating with the partner ($B = 0.607$, $p = 0.001$). Perceived strategic self-interest and prosocial motivations of the partner were not related to the child's cooperation. We also found that with increase in age children attributed interpersonal motivations to their partner more often ($B = 0.889$, $p < 0.001$) and that girls attributed more strategic self-interest to their partners than boys ($B = 1.232$, $p = 0.006$).

Table 3. Relation between perceived partner's motivations and children's decisions in the RPD

	B	*SE*	OR	95% CI	
Intercept	-3.526	1.063**			
Age	0.263	0.104*	1.30	1.06	1.60
Gender	0.106	0.168	1.11	0.80	1.55
Interpersonal motivations	0.607	0.188**	1.83	1.27	2.65
Strategic self-interest	-0.077	0.243	0.93	0.58	1.49
Prosocial motivations	0.159	0.183	1.17	0.82	1.68
Log Likelihood	-3174.652				
# Observations	5010				
# Children/Clusters	167				
χ2(df)	25.8(5)				
Pseudo R2	0.03				

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Goals and decisions in the PD game

To capture additional goals and motivations for the game, which children might have not offered spontaneously, we also asked children to identify their "thoughts during the game" from a pre-defined list of goals. Children could check multiple goals from the following list: maximizing one's own gain (I wanted to win as many points as possible), loss avoidance (I wanted to win points every round), competition (I wanted to win more points than the other kid), prosociality (I wanted to help the other kid win points), and reciprocity (I wanted to punish or reward the other kid for their choices). Wanting to win as many total points as possible was the most popular goal that children adopted for the

game (68.07%) and 73.65% of the children selected multiple goals for the game (see Table 4). Out of the children who said they played randomly, only 5 selected no other goals.

Table 4. Children's goals for the RPD

Goal selected	***N***	**% Children**
I wanted to win as many total points as possible	113	68.07%
I wanted to win points every round	85	51.20%
I wanted to win more points than the other kid	68	40.72%
I wanted to help the other kid win points	76	45.50%
I wanted to punish or reward the other kid for their choices	40	23.95%
I just pushed the buttons randomly	31	18.56%
Children who selected more than one goal	123	73.65%

To investigate the relation between children's goals for the game and their actual decisions in the game we used another nested logistic regression model, with decisions in the game as the dependent variable and children's goals as independent variables (see Table 5). Children who wanted to win more points than their partner were significantly less likely to cooperate ($B = -0.711$, $p < 0.001$). Children were more likely to cooperate if they wanted to help their partner win points ($B = 0.898$, $p < 0.001$) and if they wanted to punish or reward the other child for their choices ($B = 0.699$, $p < 0.001$).

Table 5. Relation between children's goals and decisions in the RPD

	B	*SE*	OR	95% CI	
Intercept	-2.845	0.974**			
Age	0.209	0.092*	1.23	1.03	1.48
Gender	0.061	0.149	1.06	0.79	1.42
win as many points as possible	-0.096	0.153	0.91	0.67	1.22
win points every round	-0.271	0.159	0.76	0.56	1.04
win more points than the other kid	-0.711	0.171***	0.49	0.35	0.69
help the other kid win points	0.898	0.148***	2.45	1.84	3.28
punish or reward the other kid	0.699	0.141***	2.01	1.53	2.65
push buttons randomly	-0.169	0.166	0.84	0.61	1.17
Log Likelihood	-2931.433				
# Observations	4980				
# Children/Clusters	166				
$\chi 2$(df)	93.97(8)				
Pseudo R2	0.10				

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Standard game theoretic strategies and children's RPD play

In addition to children's verbal descriptions of their play we also analyzed children's patterns of play in relation to three standard game theoretic strategies: TFT, WSLS and GRIM. First we investigated whether the game theoretic strategies would explain children's decisions in the game. For each round we determined what a pure TFT, WSLS or GRIM player would do, cooperate (coded as 1) or defect (coded as 0), given the partner's decisions up to that round. The pure TFT cooperates on the first round and then responds in kind to the partner's last move. The pure WSLS repeats their prior decision

(cooperate or defect) as long as they are "winning" (gaining 3 or 4 points) and changes their decision if they "lose" (gaining 1 or 0 points). The pure GRIM cooperates until the partner defects and then defects for the remainder of the game. We used the decisions of these pure strategies as independent variables in nested regression models to predict children's decisions, which was our dependent variable. Recall that each game consisted of 30 rounds, which were treated as repeated measures nested under each child. The slopes and intercepts were allowed to vary for each participant with clustering of standard errors.

Table 6. Standard strategies from game theory as predictors for play patterns

	Model 0	**Model 1**	**Model 2**	**Model 3**	**Model 4**
	B (SE)	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Intercept	-4.29(0.98)***	-5.02(0.98)***	-4.67(1.00)***	-4.60(0.97)***	-5.44(0.96)***
Age	0.37(0.10)***	0.36(0.09)***	0.36 (0.10)***	0.35(0.10)***	0.34(0.10)***
Gender	0.10(0.98)***	0.09(0.16)	0.10(0.17)	0.11(0.17)	0.11(0.16)
TFT		1.28(0.11)***			1.20(0.13)***
WSLS			0.65(0.08)***		0.80(0.10)***
GRIM				1.34(0.10)***	0.49(0.10)***
N	167	167	167	167	167
LL	-3214.19	-3007.15	-3160.24	-3008.11	-2884.25
Pseudo R2	0.016	0.079	0.033	0.079	0.117
AIC	6434.38	6022.29	6328.49	6024.22	5780.51
BIC	6443.73	6034.77	6340.96	6036.69	5799.21

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

All standard game theoretic strategies significantly predicted children's decisions (TFT: $B = 1.201$, $p < 0.001$; WSLS: $B = 0.801$, $p < 0.001$; GRIM: $B = 0.494$, $p < 0.001$),

which means children were more likely to cooperate on those rounds on which each of the standard strategies would cooperate. Moreover the model fit was improved by introducing the standard strategies as predictors (see Table 6). Model 1 is the null model with age and gender as covariates. The model fit is improved in Models 2, 3 and 4 that have the TFT, WSLS and GRIM strategies respectively as independent variables. Finally, Model 4, when all the strategies are introduced as independent variables, shows the best model fit. The model best fitting the data included all three standard strategies as predictors. This indicates that children played the RPD strategically, significantly approximating standard strategies from game theory.

We also wanted to estimate how well children's patterns of play matched these standard strategies and whether closeness to a certain strategy was dependent on the child's age and gender. We computed to what extent children's patterns of play matched each of the three standard strategies by comparing, for each round, the child's decision to the decisions generated by the strategy. When the child's decision was identical to that of the strategy, the match to that strategy, for that round, was coded as 1, when their decision did not match, it was coded as 0. Decisions in the game matched the TFT strategy on 62% of the rounds for all children combined, the WSLS strategy on 54% of the rounds and the GRIM strategy 69% of the rounds. 10.1 % of children adopted a perfect ALL-D strategy and defected throughout the entire game.

To investigate the influence of age and gender on the extent to which children's play patterns approximated classical strategies in the PD game, we used nested logistic regression models. Age and gender were the predictors in the models. We constructed

one model for each standard strategy tested. The dependent variable was the child's match to the strategy (coded as 1 for match, 0 for no match) on each round. Age was a significant predictor of TFT strategy match, with play patterns getting closer to the TFT strategy with increased age ($B = 0.235$, $p < 0.001$). Age and gender did not significantly predict the extent to which play patterns matched the WSLS and GRIM strategies.

Discussion

The results confirmed our first two hypotheses that a) children understood the RPD as a real social interaction and b) children explicitly engaged in strategic thinking during the game. Children's open-ended descriptions of their thoughts during the game revealed that they not only had clear ideas about their own motivations and goals but that they also formed beliefs about the partner's motivations and goals. This is striking given the minimal nature of the interaction. Most strategic games, including the RPD version designed for this study, are played online and the partners remain anonymous. To minimize noise in the data, the only means of interacting are through the decisions in the game. Thus the child never sees the partner, is never told anything about who the partner is besides that it is another child like them, and never gets to communicate with the partner in any other way. In fact in our study children were led to believe they were playing another child online, while they were actually playing against pre-programmed algorithms. Despite the minimal information provided, children generated rich descriptions of the partner's motivations and their own strategies for playing. These descriptive results provide an important validation of the RPD as a rigorous, easily

quantifiable measure of social interactions through which complex social phenomena like trust, reciprocity, fairness, etc. can be studied with children.

The study also shows that children's stated motivations were linked to their behaviors in the game. Dreber et al. (2014) found that adults' cooperation in repeated games was primarily motivated by payoff maximization in the long run. Children too, like adults, had the goal of maximizing their gain: 68% of the children endorsed the statement: "I wanted to win as many total points as possible." However this goal did not predict their level of cooperation in the game. In fact, focusing on strategic self-interest was related to decreased cooperation. By contrast, interpersonal motivations (attending to the partner and responding to the partner's decisions) was positively related to cooperation in the RPD. Also, children cooperated more when they perceived the partner's motivations as being interpersonal.

This distinction between adults and children could indicate that, overall, children lacked the intuition that sustained cooperation between partners would be beneficial for maximizing point gain in the long run. Indeed, this interpretation is supported by the fact that only a small minority of children spontaneously mentioned joint gain as a motivation for their decisions. The way children perceived their partner's motivations in the game is also telling – they rarely mentioned personal gain as a motivation for partner cooperation.

It is intriguing however to note that older, but not younger, children were more likely to mention interpersonal motivations and they were also more likely to cooperate in the RPD. It is possible that older children view interpersonal motivations, being mindful of the partner's decisions, as a means to winning more points in the long run.

This interpretation aligns, at least qualitatively, with the increased use of a TFT strategy with age.

One possibility why children were overall less likely to think of cooperation as a way to increasing personal gain might be related to the way the game was set up. In our game cooperation was achieved by the act of delivering three coins to the partner by pushing. This could be construed as an act of giving. Defection was achieved by the act of pulling one coin towards oneself which could be construed as taking. It is more counter-intuitive to think that one would maximize their gain by giving something away. Previous research has also shown that the framing of strategic games (for example in terms of giving and taking or keeping and leaving) can impact decisions and strategies in the game (see review by Kopelman et al., 2002).

Our last hypothesis concerned children's use of game-theoretic strategies in the RPD. Most children engaged in both cooperation and defection in the RPD in agreement with game theoretical predictions for repeated games (Dal Bó & Fréchette, 2011). Only 10.1 % of children defected throughout the entire game, similar to results from adult studies (Fudenberg, Rand, & Dreber, 2012). Decisions for the remainder of the children were predicted by three "pure" strategies: TFT, WSLS and GRIM. This suggests that when making new decisions in the game children take into account the history of the interaction. Standard game-theoretic strategies specify which decision should be made given the history of the game. For example, cooperation in TFT depends on the partner's decision in the previous round, cooperation in WSLS depends on the payoff from the previous round, and cooperation in GRIM depends on the partner's cooperation up to the

first defection. Overall, children's play strategies were closest to GRIM, even closer than ALL-D, (match to GRIM: 69% of rounds, match to ALL-D: 64.3%), suggesting that children are responsive to partner defection. Of the more cooperative standard strategies, children were closer to TFT than WSLS. Also children's patterns of play approximated TFT significantly more closely as a function of age: older children's play resembled TFT more. TFT is a responsive and reciprocal strategy – the player starts by cooperating and then responds in kind to the partner's decision in the previous round. This is in agreement with our finding that older children report more interpersonal motivations and thus more attention to partner's decisions in the game.

There are several limitations to this study that should be addressed in future research. First, we asked about motivations for decisions in the game only after the children had played against three different partners. The motivations might have been different for each individual partner. Future studies should investigate how motivations in the game might be affected by different kinds of partners: generally cooperative partners, generally selfish partners, forgiving partners, lenient partners, contingent responders etc. The perceived motivations of the partner are also likely to differ based on the partner's pattern of play.

Second, we did not ask about motivations in specific situations: following partner cooperation and defection, following mutual cooperation, mutual defection or exploitation of or by the partner. Again, motivations might differ based on the different circumstances and this might be relevant for how children construe the interaction in these different situations. Third, patterns of play in the game should be compared to other

versions of standard game theoretic strategies, like 2TFT in which the player defects twice in response to a defection from the partner (see an extended list in Fudenberg et al., 2012).

Finally, different factors such as background and ethnicity might influence children's interpretations of other player's actions. For example, children from minority groups may interpret the other player's actions as being more hostile. It is thus important to understand what external factors might be influencing children's motivations in the game.

In summary, this study makes several contributions to the literature. First, it shows that children construe the RPD as a social interaction. Second, it shows that children's motivations in the game are related to sophisticated social cognitions and that these motivations influence their decisions in the game. Third, our study shows that children's decisions in the game also depend on the perceived motivations of the partner: perceived interpersonal motivations of the partner increase the likelihood of cooperation. Lastly, the study shows that children play in strategic ways according to game theory and that both interpersonal motivations and more responsive strategies (more similar to TFT) increase with age.

CHAPTER 4

RESULTS: SIGNATURES OF AGGRESSION IN PRISONER'S DILEMMA PLAY

Strategic games have been recognized as useful tools for studying behavioral profiles and psychopathologies (King-Casas & Chiu, 2012). However, reactive and proactive forms of aggression have been primarily researched using questionnaire measures. Questionnaire-based measures are useful for identifying children with reactive and proactive traits, and these are indeed predictive of child outcomes. Child questionnaires based on responses to vignettes, as in the SIP model, have also proven useful in identifying the thought processes that might lead to aggressive responses. Yet these kinds of measures are also limited in that they do not allow us to study aggression as it occurs.

Some researchers have tried to rectify this issue by using computerized games that create provocation situations and allow measurement of the responses of aggressive individuals. These approaches, reviewed in Chapter 1, have been designed to be purely competitive, allowing only for negative inputs from the partner, and in some games input from the partner is received only when losing (e.g., Atkins et al., 2001; Hubbard et al., 2002; Muñoz et al., 2008; Phillips & Lochman, 2003; Waschbusch et al., 2002). Real social interactions involve an ongoing exchange of social signals and typically have an alternative outcome that is not competitive. These characteristics make children's

potentially aggressive social interactions more like cooperative dilemmas than like competitions in which there can only be one winner.

The RPD captures this characteristic of social interactions and allows for measurement of both negative (defection) and positive (cooperation) behaviors. The RPD can be construed as both a collaborative and competitive task, since outcomes depend on the joint decisions of the two play partners and incentives are set up such that exploitation of the partner leads to a maximum payoff for the round, but in the long run mutual cooperation can lead to more mutual gain. Also, throughout the game, the signal from the interaction partner is ongoing and can be either positive or negative. This context allowed us to create semi-controlled interactions that were likely to generate an aggressive (defection) response.

In this study, we used a version of the RPD game (described in detail in Chapter 2) created based on a visual interface that has been used successfully with children (Blake et al., 2015). Children played ten rounds of the game with three different partners who they believed were other children. However, the partners were pre-programmed to create combinations of play that were meant to provoke and in some cases appease. All children first played a purely-reactive TFT partner to determine their baseline rate of cooperation. Recall that this strategy cooperates in the first round and then mirrors the child's move in the prior round. Given this strategy, defection will only occur if the child defects first. Next children played an 80% cooperating (80C) and an 80% defecting (80D) partner, order counterbalanced between subjects. These partners followed a fixed pattern of play, changing from C or D in the 3rd and 7th rounds to capture children's reactions. In

addition, we measured children's reactive and proactive aggression through parent report on a standard aggression measure (PRPA, Kempes et al., 2006). We were unable to obtain parent report of aggression for some children, and some answers on the reports were incomplete. Complete data was obtained for 153 children for reactive aggression and 158 children for proactive aggression, and for 152 children for both reactive and proactive aggression. Children with missing proactive or reactive scores were excluded from analyses involving those scores. As described in Chapter 2, we used a median split to create high reactive ($N = 58$) and low reactive ($N = 95$), and high proactive ($N = 33$) and low proactive ($N = 125$) groups respectively.

Previous research has shown that children with high reactive aggression tend to interpret ambiguous social interactions, such as someone bumping into you, as hostile in nature (Dodge & Coie, 1987; Fontaine & Dodge, 2006). Reactive children also tend to "hold a grudge" longer than children without aggressive tendencies (Waschbusch et al., 2002). These differences from control groups suggest that the negative interpretation of social behavior and the failure to forgive negative actions may be the drivers of aggressive behaviors in reactive aggressive children. We thus hypothesized that:

a) high reactive aggressive children would react more strongly to defection in the RPD, demonstrated by lower cooperation immediately following partner defection, as compared to children low in reactive aggression,

b) high reactive aggressive children would be less forgiving of partner defection, demonstrated by a longer string of defections in the rounds following partner defection, compared to children low in reactive aggression,

c) the predicted, unforgiving response for (b) is a characteristic of the GRIM strategy described in Chapter 3; thus, we also assessed whether high reactive children were more likely to use this strategy in a separate strategy analysis.

Proactive aggression has been characterized by calculated, instrumental behavior (Crick & Dodge, 1996). As such, high proactive aggressive children would more likely be motivated by payoff. We hypothesized that they would try to obtain points in the RPD game through exploitative aggressive strategies:

a) high proactive aggressive children would seek an advantage by defecting more in response to partner cooperation, compared to low proactive children; this move would be an attempt to create a DC outcome with highest payout to the child,

b) high proactive aggressive children would defect more in the last round,

c) high proactive aggressive children would cooperate more after mutual defection; this move would be an attempt to return to a combination of decisions (CC) that would yield a higher payout to the child,

d) high proactive aggressive children would be more likely to use a WSLS strategy, which is a more exploitative, gain-focused and less impulsive strategy, compared to low proactive children.

Note that the prediction for (c) differs from (a) but is in agreement with (b) in how children would respond to mutual cooperation (CC). If children actively seek to exploit the partner, they would defect after CC, but a WSLS strategy would cooperate since the child earns a high amount of points. Also, WSLS strategy is consistent with (b) and would cooperate after mutual defection, since DD produces a "losing" payout. Because

nothing is known about how proactively aggressive children might approach the game, we tested both possibilities in different analyses.

Data Analysis Plan

We analyzed the relation between aggression profiles (low or high reactive or proactive aggression groups) and RPD game play using a series of nested logistic regression models. We used logistic models because the RPD produces a binary decision in each round: cooperate (coded as 1) or defect (coded as 0). Each child played thirty rounds in total, which were treated as a repeated measure with rounds nested under each participant. The slopes and intercepts were allowed to vary for each participant with clustering of standard errors. Unlike studies that paired participants randomly with each other, we standardized the experience of all of our participants by having them play with pre-programmed algorithms. Since their strategies were part of the experimental design, we introduced the partner as a categorical predictor in some of the relevant logistic models.

The analyses followed a standard approach for RPDs (Blake et al., 2015; Fudenberg et al., 2012). We first tested overall cooperation across all rounds and across partners comparing the high and low aggression groups for the reactive and proactive measures. We then systematically analyzed decisions in particular circumstances to better understand the behavioral patterns associated with each profile:

a) following partner decision in the previous round (C or D)

b) following different child-partner decision combinations (CC, CD, DC, DD)

c) and in the round-by-round responses to the 80C and 80D partners

Finally we tested whether children's strategies in the game matched some of the standard game theoretic strategies for the RPD (TFT, WSLS and GRIM) and whether the match predicted children's aggression profiles. We calculated the match to these strategies in the same way as explained in Chapter 3.

Results

RPD decisions and aggression profiles

We first examined the effect of testing location (online vs. schools) on decisions in the RPD game. We also checked whether location predicted belonging to the low or high reactive and proactive aggression groups. No differences were found and thus data was combined for further analyses.

To determine whether reactive or proactive aggression (low or high) predicted the probability of cooperation in the RPD, we created a nested logistic regression model with RPD decisions from 30 rounds, nested under each participant, as the dependent variable (coded 1 for C and 0 for D). Age (in years) and gender were introduced as covariates (see Table 7). Older children cooperated more ($B = 0.312$, $p = 0.002$) and children in the high reactive aggression group cooperated significantly less (see Figure 6a) than children in the low reactive aggression group ($B = -0.447$, $p = 0.022$). There were no differences between children in the low and high proactive aggression groups in terms of cooperation. Gender did not significantly predict cooperation either.

Table 7. Probability of cooperation by low and high reactive and proactive aggressive children

	B	*SE*	95% CI	
Intercept	-3.606	1.03***		
Age	0.312	0.10**	0.11	0.51
Gender	0.049	0.17	-0.29	0.39
Reactive Aggression (Low/High)	-0.447	0.19*	-0.83	-0.06
Proactive Aggression (Low/High)	0.205	0.21	-0.20	0.61
Log Likelihood	-2930.18			
# Observations	4560			
# Children/Clusters	152			
$\chi 2$(df)	15.07(4)			
Pseudo R2	0.02			

* p < 0.05, ** p < 0.01, *** p < 0.001

Contingent Cooperation

The RPD is an interactive task, and decisions are often made in response to the partner. Following prior RPD studies, we next investigated how children reacted to the partner's decisions in the previous round (see Table 8). We first confirmed that all children engaged in contingent play: they cooperated more following partner cooperation than following partner defection ($B = 1.257$, $p < 0.001$). Older children cooperated more following partner cooperation than younger ones ($B = 0.467$, $p < 0.001$). Consistent with prior RPDs with children, participants in this sample did not play the RPD randomly but were responding to partner decisions (see Figure 6b).

Figure 6. Overall cooperation and contingent cooperation in the RPD for children in the low and high reactive aggression groups

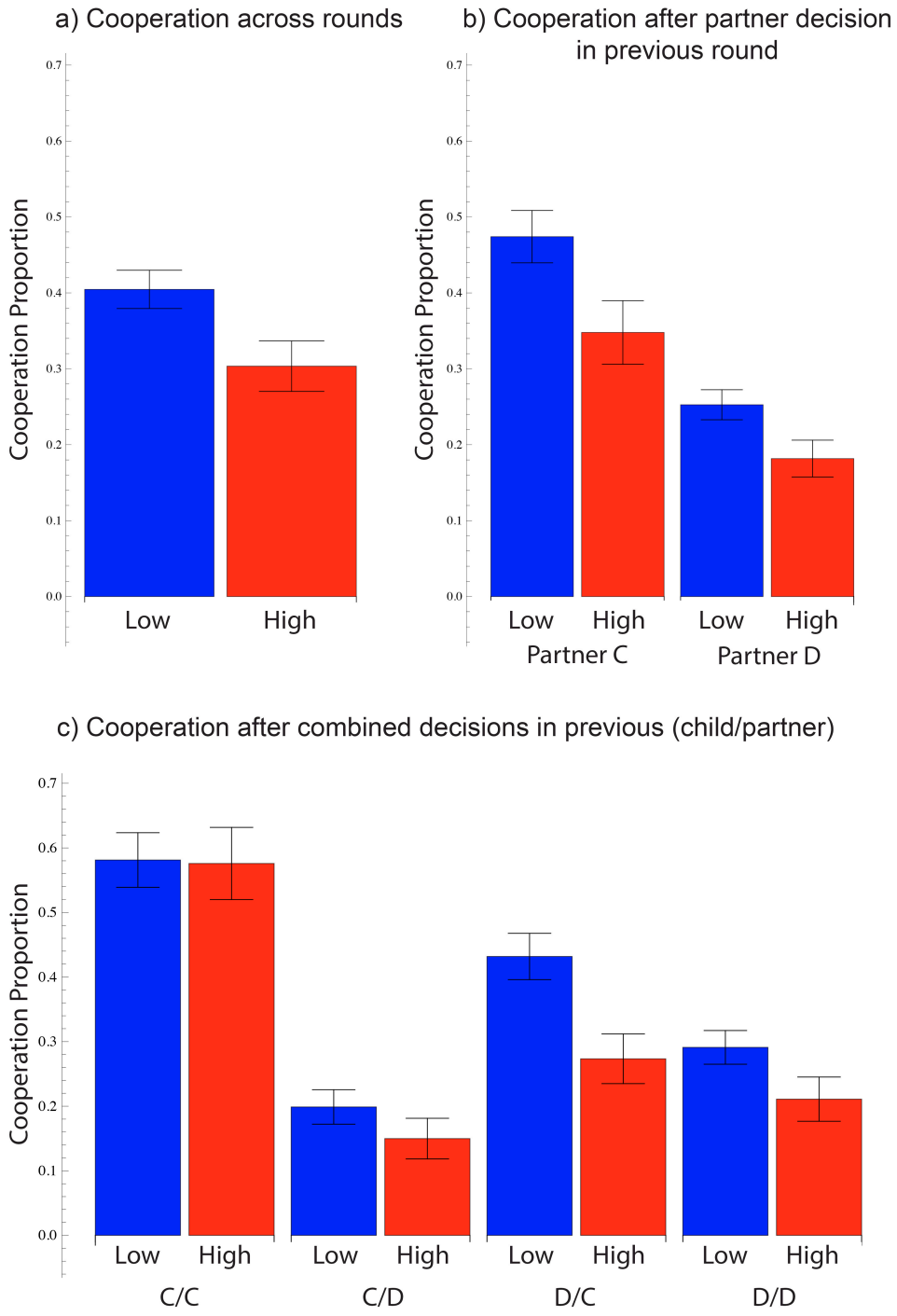


Table 8. Probability of cooperating after partner C and partner D for low and high reactive and proactive aggressive children

	B	*SE*	95% CI		*B*	*SE*	95% CI	
Intercept	-4.522	1.342**			-1.622	1.075		
Age	0.454	0.131**	0.20	0.71	0.035	0.109	-0.18	0.25
Gender	0.050	0.220	-0.38	0.48	0.072	0.177	-0.27	0.42
Reactive	-0.457	0.233†	-0.91	0.00	-0.563	0.196**	-0.95	-0.18
Proactive (Low/High)	0.042	0.276	-0.50	0.58	0.473	0.195*	0.09	0.86
Log Likelihood	-1393.29				-1020.47			
# Observations	2085				2019			
# Children/Clusters	152				152			
χ2(df)	16.37(4)				12.42(4)			
Pseudo R2	0.03				0.01			

† p=0.05 * p < 0.05, ** p < 0.01, *** p < 0.001

Comparing the high and low reactive aggression groups revealed that cooperation was lower for high reactive children both following partner cooperation, B = -0.457, p = 0.05, 95% CI [-0.91, 0.00], and following partner defection, B = -0.563, p = 0.004, 95% CI [-0.95, -0.18]. Surprisingly, high proactive aggressive children were more likely to cooperate than low proactive aggressive ones following partner defection (B = 0.473, p = 0.015). There were no differences in cooperation between high and low proactive aggressive children following partner cooperation.

The differences in contingent cooperation between the high and low aggression groups may have depended on the child's decision in the prior round in addition to the partner's decision, as has been found in prior work (Blake et al., 2015). To examine this possibility, we next analyzed the effect of the combined (child, partner) decisions in the

prior round on cooperation in the current round (see Table 9 and Figure 6c). Since each player has a binary choice between C and D, four decision combinations are possible: mutual cooperation (CC), child cooperation and partner defection (CD), child defection and partner cooperation (DC) and mutual defection (DD).

Recall that mutual cooperation (CC) produces the highest total payoff for the dyad (3 points each) and thus represents a positive and successful interaction. Following mutual cooperation children tended to cooperate again. Holding proactive aggression constant, there were no differences between the low reactive and high reactive groups (low reactive: 58.12% C, high reactive: 57.58% C). Holding reactive aggression constant, there were no significant differences between high and low proactive aggression groups either (low proactive: 59.92% C, high proactive: 49.28% C). Older children tended to cooperate more after mutual cooperation ($B = 0.417$, $p = 0.019$).

By contrast, a CD interaction results in the lowest payoff for the child (0 points) and the highest for the partner (4 points). Following these cases of being exploited by the partner, all children tended to defect. There were no differences between reactive aggressive groups (low reactive: 19.88% C, high reactive: 14.98% C) or proactive aggressive groups (low proactive: 18.64% C, high proactive: 13.64% C). Thus, when the child cooperated in the prior round, decisions depended on what the partner did in the prior round but not on the level of aggression. Older children tended to cooperate more after being exploited by the partner ($B = 0.414$, $p = 0.024$) and girls tended to cooperate less than boys ($B = -0.627$, $p = 0.031$).

Table 9. Probability of cooperation after combined decisions in the previous round

	CC				**CD**			
	B	*SE*	95% CI		*B*	*SE*	95% CI	
Intercept	-3.219	1.813			-5.049	1.899**		
Age	0.417	0.177*	0.07	0.76	0.414	0.183*	0.05	0.77
Gender	-0.031	0.281	-0.58	0.52	-0.627	0.291*	-1.20	-0.06
Reactive Aggression	-0.036	0.306	-0.63	0.56	-0.130	0.332	-0.78	0.52
Proactive Aggression (Low/High)	-0.291	0.360	-0.10	0.41	-0.237	0.437	-1.09	0.62
Log Likelihood	-523.60				-315.52			
# Observations	896				611			
# Kids/Clusters	122				134			
χ2(df)	6.65(4)				12.27(4)			
Pseudo R2	0.02				0.04			

	DC				**DD**			
	B	*SE*	95% CI		*B*	*SE*	95% CI	
Intercept	-3.534	1.490*			0.521	1.394		
Age	0.287	0.147	0.00	0.57	-0.208	0.141	-0.48	0.07
Gender	0.137	0.251	-0.35	0.63	0.474	0.216*	0.05	0.90
Reactive Aggression	-0.702	0.260**	-1.21	-0.19	-0.692	0.230**	-1.14	-0.24
Proactive Aggression (Low/High)	0.096	0.309	-0.51	0.70	0.762	0.224**	0.32	1.20
Log Likelihood	-706.29				-677.63			
# Observations	1189				1408			
# Children/Clusters	152				152			
χ2(df)	11.69(4)				26.50(4)			
Pseudo R2	0.03				0.04			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Differences between reactive aggression groups emerged when the child had defected in the prior round. When the child exploited the partner (DC) in the prior round, children low in reactive aggression were more likely to cooperate (43.2% C) compared to

those high in reactive aggression (27.3 % C, B = -0.702 p = 0.007), controlling for proactive aggression. Given that the children in these cases had just received the highest possible payoff (4 points), it is notable that the children low in reactive aggression seemed to shift towards re-establishing cooperation, while children high in reactive aggression continued to defect. There were no significant differences between children low in proactive aggression and children high in proactive aggression (low proactive: 38.05% C, high proactive: 30.79% C) following DC rounds. It is surprising that children high in proactive aggression did not continue defecting after receiving the highest possible payoff of 4 points.

For both reactive and proactive aggression a difference between low and high groups emerged after mutual defection (DD), but in the opposite direction. In a DD round, both children receive a low payoff of 1 point each. Children high in reactive aggression were less likely to cooperate on the next round compared to children low in reactive aggression (high reactive: 21.09% C, low reactive: 29.10% C, B = -0.692 p = 0.003), controlling for proactive aggression. On the contrary, children high in proactive aggression were significantly more likely to cooperate on the next round compared to children low in proactive aggression (high proactive: 30.46% C, low proactive: 24.52% C, B = 0.762, p = 0.001), while keeping reactive aggression constant. This suggests that children high in proactive aggression might have understood that mutual defection can have a detrimental effect on the payoff, and were trying to restore partner cooperation. Girls were more likely to cooperate following mutual defection (B = 0.474, p = 0.028).

Combined, these responses suggest that children with a low reactive aggression profile are more likely to return to cooperation after they defected. However, their return to cooperation after defection is modulated by the partner's decision, because low reactive aggressive children but not high reactive aggressive children cooperate more after DC than after DD ($B = 0.623$, $p < 0.001$). Also, high proactive aggressive children are more likely to re-establish cooperation but only after mutual defection (DD).

In summary, these results replicate and extend prior work linking children's aggression to contingent play in the RPD. One prior study found that children rated high on conduct problems using a general scale of externalizing behavior cooperated less after the they had defected in the prior round (Blake et al., 2015). We find similar results for children rated high on reactive aggression compared to low reactive aggression but find no differences in RPD play based on proactive aggression profile. However, given that the current experiment used pre-defined partner strategies, as opposed to live play with a peer, we can additionally examine how children with different aggression profiles respond to each of the three partners and compare responses to particular rounds of play designed to provoke an aggressive response.

Cooperation with different partners and aggression profiles

We first tested children's cooperation on the first round of play with each partner. Cooperation on the first round did not differ between high and low reactive and proactive aggression groups for none of the three partners (TFT, 80C and 80D). Also, the order of the 80C and 80D partners (which was counterbalanced) did not significantly predict

cooperation in the first round of play with these partners. Thus, children approached each partner as distinct and their experience with a mostly cooperating or defecting partner did not spill over into the initial interaction with the new partner.

Each child first played with a TFT partner that cooperated on the first round and then mirrored the child's decision on the prior round. The TFT partner offers insight into children's baseline approach to the task. In particular, given that the TFT partner starts with cooperation and will continue to cooperate if the child does so on the first round, the child must be the first to defect. We first examined whether children defected at all with this partner, and found that 81.1% of children defected at least once during play with the TFT partner and 44.0% of the children defected in the first round. There were no overall differences in cooperation level between high and low reactive or high and low proactive groups.

The next two partners, 80C and 80D, were designed to elicit responses to particular patterns of play that were meant to provoke or appease. Both partners followed a fixed pattern of play: 80C always defects on rounds 3 and 7 and cooperates otherwise, and 80D always cooperates on rounds 3 and 7 and defects otherwise. The order of these two partners was counterbalanced and no order effects were found. With the 80C partner children high in reactive aggression cooperated less than children low in reactive aggression (B = -0.571, p = 0.009). There were no differences between high and low proactive aggressive groups in the level of cooperation with 80C. With the 80D partner, again children high in reactive aggression cooperated less than children low in reactive aggression (B = -0.591, p = 0.006). There were again no differences between high and

low proactive aggressive groups in cooperation with 80D. Because we found no differences between the high and low proactive aggression groups, the follow-up analyses presented below refer only to reactive aggression.

Reactive aggression and responses to fixed strategy partners

We next examined play patterns with each of the 80C and 80D partners in order to test how children with low and high reactive aggression responded after the partner deviated from the overall pattern of cooperating or defecting (i.e.,, after round 3 and 7; see Figure 7). We analyzed decisions over the course of the interaction with the partner by grouping rounds into three logical periods (see Table 9). The first three rounds represented a *baseline* during which children adjusted their cooperation to each particular partner (following Muñoz et al., 2008). Given that the 80C and 80D partners deviated from their general pattern of cooperation and defection in rounds 3 and 7, rounds 4 and 8 represented the child's *response* to that change. For the 80C partner in particular, we predicted that children high in reactive aggression would be more likely to defect in the response rounds, showing a hostile attribution bias to the partner's defection. The rounds following the response rounds (5,6,7,9,10) represented *recovery* periods during which children re-adapt to the partner. This grouping of rounds follows prior work that found a faster "dissipation" of anger in control groups compared to reactive children (Waschbusch et al, 2002; Helseth et al., 2015). For each grouping, we combined children's decisions (treating decisions as repeated measures by nesting rounds under each participant) and compared the high and low reactive aggression groups.

The 80C partner

As mentioned above, with the 80C partner overall, children in the low reactive aggression group were more likely to cooperate than children in the high reactive group, unlike when playing against the TFT partner. However, the degree of cooperation varied systematically over rounds in response to the partner's decisions. For the baseline rounds, there was no difference between low reactive and high reactive aggression children in the probability of cooperation in these rounds. In the response rounds, there were again no differences between the high and low reactive aggression groups, suggesting that both groups of children responded to defection with defection. In the recovery rounds, children in the low reactive aggression group were significantly more likely to return to cooperation than children in the high reactive aggression group (B = -0.711, p = 0.004). This was also true when rounds 5 and 9 were taken alone (B = -0.714, p = 0.016). Combined these results suggest that both high and low reactive children behave similarly in their initial interactions with a generally cooperative partner and both punish the partner's defection. However, high and low reactive aggression children differ in how they recover from the partner's defection, with low reactive aggressive children more likely to re-establish cooperation.

Figure 7. Cooperation for low and high reactive aggression groups by round with 80C and 80D partners

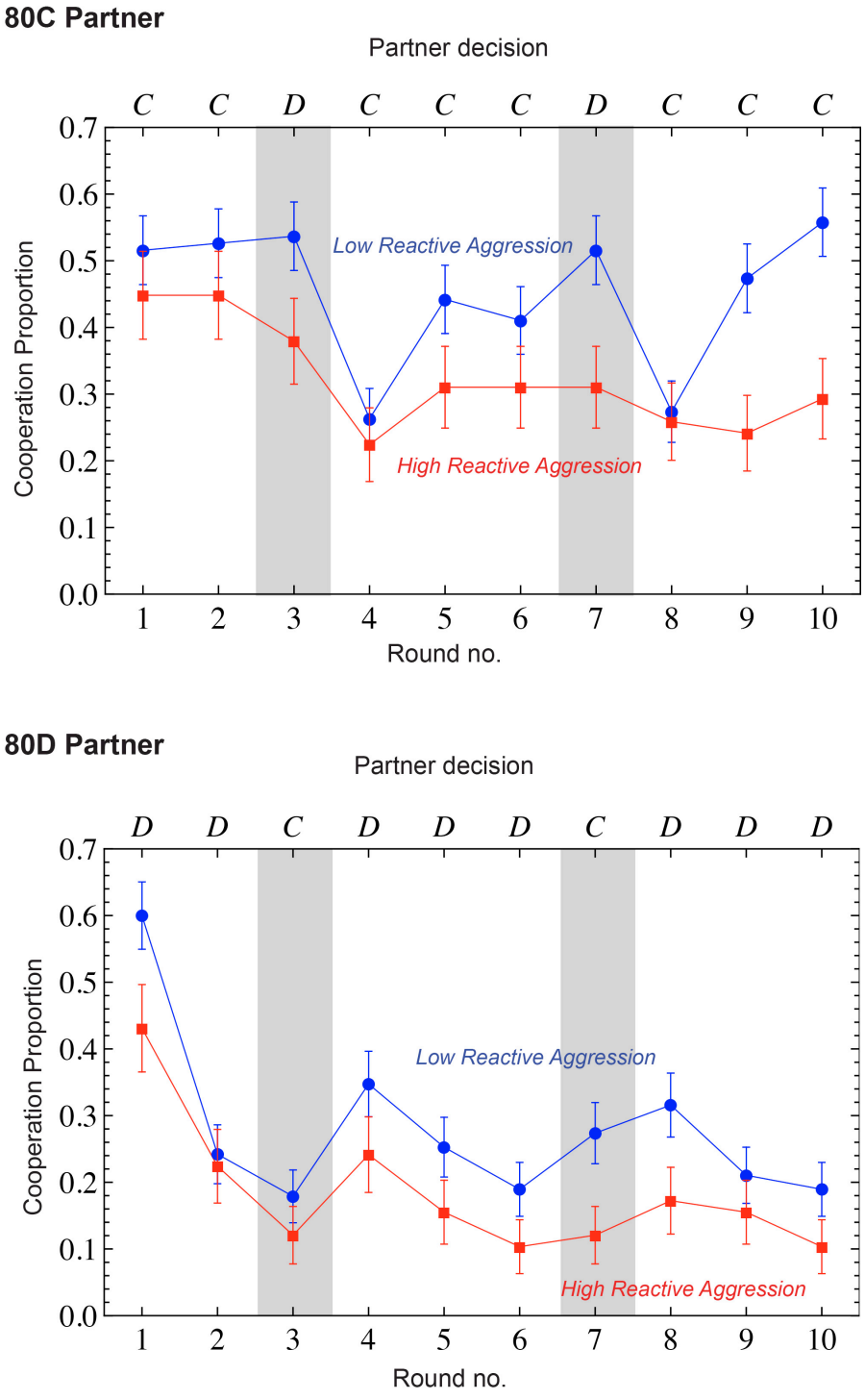


Table 10. Regression models for cooperation for low and high reactive aggression groups with 80C and 80D partners

80C	**Round 1**		**Rounds 2, 3**		**Rounds 4, 8**		**Rounds 5-7, 9-10**	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	-3.886	2.076	-3.093	1.809	-2.233	1.956	-4.603	1.572**
Age	0.387	0.204	0.329	0.178	0.119	0.188	0.437	0.152**
Gender	0.068	0.331	-0.204	0.297	0.045	0.293	0.204	0.242
Aggression (Low/High)	-0.188	0.341	-0.438	0.308	-0.113	0.311	-0.711	0.245**
Log Likelihood	-103.84		-206.91		-174.35		-492.73	
# Observations	153		306		306		765	
# Children/Clusters	153		153		153		153	
χ2(df)	4.24(3)		6.14		0.75		17.43	
Pseudo R2	0.02		0.02		0.002		0.05	

80D	**Round 1**		**Rounds 2, 3**		**Rounds 4, 8**		**Rounds 5-7, 9-10**	
	B	*SE*	*B*	*SE*	*B*	*SE*	*B*	*SE*
Intercept	-4.744	2.134*	-1.823	1.712	-4.981	1.621**	-1.537	1.504
Age	0.524	0.209*	0.033	0.172	0.428	0.158**	0.034	0.148
Gender	-0.242	0.341	0.311	0.292	-0.121	0.294	-0.108	0.243
Aggression (Low/High)	-0.631	0.348	-0.206	0.308	-0.584	0.316	-0.680	0.272*
Log Likelihood	-100.04		-150.52		-176.15		-362.67	
# Observations	153		306		306		765	
# Children/Clusters	153		153		153		153	
χ2(df)	10.78		2.13		10.71		6.67(3)	
Pseudo R2	0.05		0.01		0.03		0.01	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The 80D partner

For the 80D partner we again grouped the rounds into baseline, response and recovery and compared the cooperation probabilities for the reactive aggression profiles. Baseline rounds 2 and 3 represent a reaction to two consecutive defections on the part of the partner. We found no difference between the low and high reactive aggression groups in the probability with which they engaged in cooperation in these rounds (see Table 10). Rounds 4 and 8 represent a response to partner cooperation in rounds 3 and 7, and measure the immediate reaction to "an olive branch" extended by a partner who established itself as generally non-cooperative. There were again no differences between low and high reactive aggression children in terms of the probability of cooperating on these rounds. Finally, rounds 5, 6, 7, 9 and 10 represent the extended reaction to a mostly non-cooperative partner that does sometimes cooperate albeit rarely. Low reactive aggressive children were significantly more likely to cooperate in these rounds than high reactive aggressive children (B = -0.680, p = 0.012). This suggests that low reactive aggressive children are more likely to seek cooperation even with a generally non-cooperative partner compared to high reactive aggressive children.

Transitions between periods for 80C and 80D

A second analysis focused on the transitions between the baseline, response and recovery periods for both the 80C and 80D partner. Children with different aggression profiles may have different responses to the partner's decisions in rounds 3 and 7 given

that these are the first and second deviations from the partner's general pattern of cooperating and defecting. We examined these rounds separately (see Table 11).

Table 11. Cooperation with partner before and after partner defection for reactive aggression profiles

80C	**Low Reactive Aggression** n=95, obs.=190				**High Reactive Aggression** n=58, obs.=116			
	B	*SE*	95% CI		*B*	*SE*	95% CI	
Round 3 to 4	-1.180	0.33***	-1.83	-0.53	-0.785	0.42	-1.61	0.04
Round 4 to 5	0.798	0.31*	0.18	1.41	0.463	0.36	-0.24	1.16
Round 7 to 8	-1.048	0.33**	-1.69	-0.40	-0.259	0.40	-1.04	0.52
Round 8 to 9	0.872	0.3**	0.29	1.46	-0.099	0.45	-0.99	-0.79
80D								
Round 3 to 4	0.905	0.35*	0.21	1.60	0.864	0.48	-0.08	1.81
Round 4 to 5	-0.472	0.35	-1.15	0.21	-0.591	0.49	-1.54	0.36
Round 7 to 8	0.203	0.26	-0.31	0.71	0.425	0.47	-0.50	1.35
Round 8 to 9	-0.553	0.34	-1.22	0.11	-0.130	0.51	-1.12	0.86

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

When playing the 80C partner, low reactive aggressive children showed a significant drop in cooperation in response to partner's defection both in round 4 (Round 3: 53.68% C vs. Round 4: 26.31% C, $B = -1.180$, $p < 0.001$) and in round 8 (Round 7: 47.36% C, $B = -1.048$, $p = 0.002$ vs. Round 8: 27.36% C) as compared to rounds 3 and 7, respectively. However, there was a subsequent significant increase in cooperation in rounds 5 and 9 as compared to 4 and 8 respectively (Round 5: 44.21% C, $B = 0.798$, $p <$

0.001; Round 9: 47.36% C, $B = 0.872$, $p = 0.003$), suggesting that low reactive aggressive children promptly recovered and returned to cooperation. In contrast, high reactive aggressive children never recovered after the first partner defection. They showed decreased cooperation in round 4 in response to partner defection (Round 3: 37.93% C vs. Round 4: 22.41% C, $B = -0.785$, $p = 0.06$, 95% CI [-1.61, 0.04]), but no subsequent significant increase in cooperation in round 5 (Round 5: 31.03% C). Further, there were no significant differences in cooperation between rounds 7 and 8, and 8 and 9 for high reactive aggressive children. Thus, high reactive children appeared to follow a GRIM strategy of defecting in response to the partner's first defection and continuing to defect thereafter.

When playing the 80D partner, children Low in Reactive Aggression showed a significant increase in cooperation in round 4 (Round 3: 12.06% C vs. Round 4: 24.13% C, $B = 0.905$, $p = 0.01$), in response to partner's first cooperative move but no subsequent changes in cooperation when transitioning between the response and recovery rounds. There were no significant differences in cooperation between transition rounds for the high reactive aggressive children.

Last round defection

A common strategy in the RPD is to defect on the last round (Andreoni, & Miller, 1993; Normann & Wallace, 2012), since the partner will not have a chance to retaliate. To explore whether children engaged in so-called "backwards induction" and switched from cooperation to defection in the last round, we compared cooperation on the last

round with cooperation on the second-to-last round. We found no significant differences in cooperation on the last round as compared to the second-to-last round for all children combined ($B = 0.085$, $p = 0.394$). This was also the case when we analyzed data separately from the high proactive ($B = -0.152$, $p = 0.585$) and high reactive aggression groups ($B = -0.135$, $p = 0.532$).

Aggression and the standard game theoretic strategies

We used the same method described in Chapter 3 to calculate how closely children's patterns of play matched standard strategies from game theory: TFT, WSLS and GRIM. We then investigated whether the extent to which children's pattern of play across all partners matched the standard PD strategies was related to their aggression profiles. TFT is a cooperative and reciprocal strategy, in which one starts by cooperating and then responding in kind to the partner. WSLS is a strategy focused on winning and unlike TFT it allows continued defection after exploiting the partner (DC). GRIM is an unforgiving strategy, which does not restore cooperation after partner's first defection.

We used a nested logistic model to predict belonging to the high or low reactive profiles (coded 1 for high, 0 for low). The independent variable was the match to each of the classical strategies (coded 1 for match, 0 for non-match) for each round across the three partners. Each round was considered a repeated measure and nested under each participant. Closer match to the GRIM strategy significantly predicted a higher chance of belonging to the high reactive aggressive group ($B = 0.425$, $p < 0.001$; see Table 12). No other strategy match significantly predicted group membership for reactive aggression. A

similar logistic model was used to investigate the relation between proactive aggression profiles and strategy matches, but none of the strategies significantly predicted proactive aggression across all partners.

Table 12. Aggression profiles and standard RPD strategies

	Reactive Aggression (0 - Low, 1 - High)				**Proactive Aggression** (0 - Low , 1 - High)			
	B	*SE*	95% CI		*B*	*SE*	95% CI	
Intercept	2.734	2.089			1.028	2.381		
Age	-0.307	0.208	-0.71	0.09	-0.241	0.237	-0.71	0.22
Gender	-0.452	0.34	-1.12	0.21	0.097	0.396	-0.68	0.87
TFT	-0.242	0.136	-0.51	0.02	-0.155	0.158	-0.46	0.15
WSLS	-0.157	0.115	-0.38	0.07	0.106	0.126	-0.14	0.35
GRIM	0.425	0.111***	0.21	0.64	0.041	0.126	-0.21	0.29
Log Likelihood	-2963.5				-2408.033			
# Observations	4590				4740			
# Kids/Clusters	153				158			
χ2(df)	18.69(5)				7.15(5)			
Pseudo R2	0.03				0.01			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Since children modulated their responses based on the partner they played, we tested whether the patterns of play with specific partners and their match to the classical PD strategies predicted children's aggression profiles. Closer match to the GRIM strategy predicted a higher probability of belonging to the high reactive aggressive group when playing against TFT ($B = 0.447$, $p = 0.009$), 80C ($B = 0.271$, $p = 0.018$) and 80D ($B = 0.505$, $p = 0.011$). Closer match to the TFT strategy predicted a lower probability of belonging to the high proactive aggression group when playing TFT ($B = -0.385$, p =

0.049) and a lower probability of belonging to the high reactive aggression group when playing against 80C (B = -0.428, p = 0.014). Lastly, closer match to the WSLS strategy predicted a higher probability of belonging to the high proactive aggressive group when playing the TFT partner (B = 0.409, p = 0.053) and a lower probability of belonging to the high reactive aggressive group (B = -0.360, p = 0.025).

Discussion

This study demonstrates that the RPD can be fruitfully used to study key aspects of both reactive and proactive aggression in children. The task allows for systematic investigations of interaction patterns that distinguish children with high and low, reactive and proactive types of aggression. We found clear signatures of reactive aggression in RPD play: decreased overall cooperation and lack of forgiveness for partner defection. Also, we found tentative indications that proactive aggression might also be related to more subtle differences in RPD strategy.

Our results partially supported our first hypothesis: that children high in reactive aggression would react more to partner defection. Both low and high reactive aggressive children defected in response to partner defection. While there were no differences in the likelihood of defection between the two groups after being exploited by the partner (child C and partner D in previous round), children high in reactive aggression cooperated less than low reactive aggressive children after mutual defection (DD in previous round). This result is consistent with findings that all children react to high provocation in games, but only reactive aggressive children react to low provocation (Muñoz et al., 2008). The CD

rounds, in which the interaction results in the highest payoff for the partner and no payoff for the child, can be considered a high provocation outcome compared to the DD rounds, in which the interaction results in a low payoff for both the child and the partner.

Our second hypothesis that high reactive aggressive children would be less forgiving of partner defection was confirmed. Following partner defection, low reactive aggressive children were willing to forgive the partner and attempt to re-establish cooperation, giving them the benefit of the doubt. By contrast, children with high reactive aggression began their interaction with each partner at a typical level of cooperation but quickly defected in response to the partner's defection and continued to defect thereafter. Notably, with the 80C partner the high reactive aggressive group did not cooperate even after multiple cooperative moves by the partner. This pattern resembles the GRIM strategy from game theory, defined as C until partner D, then D. Closer adherence to a GRIM strategy predicted membership in the high reactive aggression group both when the rounds were all combined and when they were examined for each of the partners. This result confirms our third hypothesis for reactive aggression.

These results are similar to findings by Waschbush et al. (2002) who investigated the dissipation of anger in a competitive reaction time task. The rules of the game were that each time one of the players lost, the winner could punish the loser by taking away points from them, without being able to keep them for oneself. The game was rigged such that the child lost to the partner, who punished at high levels by taking away a lot of points. The child won for the next four rounds. Waschbush et al. measured the amount of points children took from the partner in the four trials after the partner took a lot from

them. For these "anger dissipation trials," they found that children with a co-morbid diagnosis of ADHD and ODD/CD, a diagnostic profile more typical of reactive aggression (Kempes et al., 2005), continued to punish at high levels. Unlike the control group, they did not decrease point taking over the four trials following high provocation.

Although our findings are consistent with those of Waschbush et al. (2002), several differences in the tasks warrant further discussion. Unlike the competitive task used by Waschbush et al., in which there is only one winner per round, mutual cooperation in the RPD can result in a high, equal payoff for both partners, a positive and thus desirable interaction. In addition, in the competitive task the child receives an unambiguous negative message from the partner in rounds in which the partner wins, but no signal from the partner when the child wins. Thus, during the "anger dissipation" trials the partner cannot signal a willingness to return to better behavior. Taking away points from the partner during these "anger dissipation" trials is thus solely attributed to anger and amounts to punishment of the partner. In the RPD however, the signal from the partner is ongoing. Thus, when the child continues to defect while the partner cooperates, as occurs with the 80C partner, their aggressive response is resistant to overtures to re-establish cooperation from the partner.

This difference in the RPD compared to competitive tasks demonstrates the persistence of a negative response in a somewhat more realistic context, given that they can see the partner's responses each round. However, the specific motives that lead to the GRIM strategy remain unclear. One possibility is that children lose trust in the partner and choose to defect out of fear of being exploited again. This explanation gains some

support from one study using a Trust Game in which children with conduct problems were both less trustworthy and less trusting (Sharp et al., 2011). A second possibility is that children are actively trying to exploit the partner when they defect. In this case, children may initially defect out of anger, but when they receive a high payout for doing so, given that the partner cooperated, they may continue to seek this advantage. While this kind of exploitative motive is more characteristic of proactive than reactive aggression, it represents a plausible alternative. Although the current study cannot distinguish between these motives, simply asking children what they expect the partner to do before each round would provide an answer. If they defect when they expect the partner to defect, they lack trust in the partner. By contrast, if they defect when they expect the partner to cooperate, they are trying to take advantage of the partner.

A third possible explanation for the GRIM strategy is that trust and anger work together to generate a response. Reactive aggression is often associated with a history of abuse (Vitaro et al., 2006). It is possible that exposure to chronic threat can cause a lack of trust, and that initially natural defensive responses over time take the form of exaggerated displays of anger and punishment. Future experiments could disentangle the two motives by comparing the reaction to partner defection of high reactive aggressive children when the partner has the reputation of being a high-cooperator and thus of being trustworthy (the child could be told that this partner has cooperated a lot in the past with other children) compared to a partner with a reputation of being a high-defector.

For proactive aggression, our first two hypotheses concerned advantage seeking behaviors: defecting in response to partner cooperation and defecting in the last round of

the game. High proactive aggressive children did not defect more after partner cooperation. This may have been due to the fact that they viewed a mutual cooperation interaction (CC), where they each got 3 points, as a successful outcome. In addition, they may have considered the one extra point that they would receive by exploiting the partner (CD round) not worth the risk of the partner retaliating. Since proactive aggressive children are typically motivated by instrumental goals (Crick & Dodge, 1996), future experiments could test their propensity to resort to exploitation for material gain by changing the payoff matrix of the RPD and making exploitation more appealing. This would probably increase defection rates for all children, but proactive aggressive children should seek this option more often if their goal is to exploit the partner.

We also found no support for the hypothesis that high proactive aggressive children would more likely switch to defection on the last round. Although we drew attention to the last round in the games by clearly announcing it through a written message that appeared at the top of the screen, children might have not understood its significance – that there can be no retaliation from the partner following it. Blake et al. (2015) also failed to find any decrease in cooperation in the last round, a phenomenon that often appears in games with adults (Andreoni & Miller, 1993). It is possible that children simply do not reason about the game at the same level of detail and maturity, and that children's abilities to engage in backward induction are limited. However, this does not preclude the possibility that proactive aggressive children might have engaged in more exploitative last-round defection had they been made aware of its characteristics. Future experiments could in principle elucidate this.

We did find support for our third hypothesis: children high in proactive aggression were more likely to cooperate after mutual defection than children low in proactive aggression, which can be interpreted as seeking to restore cooperation and return to a higher paying decision combination, CC or DC. This is consistent with the idea that proactive aggressive children are motivated by payoff in interactions. This pattern of play of high proactive aggressive children resembles the WSLS strategy. A characteristic feature of the WSLS strategy is that one shifts to C when faced with mutual defection, because DD is considered a losing payout.

Indeed, our last hypothesis for proactive aggression concerned the use of a particular, instrumental strategy: WSLS. We did find tentative support for this prediction. When playing the TFT partner, high proactive aggressive children were more likely to follow the WSLS strategy and were less likely to use a TFT strategy. There are two key differences between the two strategies. One is in how the player responds to DC, the case in which the child defects and the partner cooperates leading to 4 points for the child and none for the partner. The WSLS strategy considers this outcome a "win" and thus "stays" with their current move and defects again. By contrast, the TFT strategy sees that the partner has cooperated and cooperates in response. Another difference is in how the player responds to DD, the case in which both partners defect on each other, leading to 1 point for each of the partners. The WSLS strategy considers this outcome a "loss" and thus "shifts" their move to cooperation. By contrast, the TFT strategy sees that the partner has defected and defects in response. Thus, the WSLS strategy will continue to exploit the partner after successfully exploiting them but will also return to cooperation if

response to a low mutual defection payout. Intriguingly, WSLS only appeared for the high proactive aggression children when playing the TFT partner. Recall that for this partner, the child must be the first to defect. By extension it seems that this group of children are likely initiating defect and then trying to exploit the partner again, but will return to cooperation if mutual defection ensues. This stands in contrast to the GRIM strategy used by the high reactive aggression children, which is purely reacting to the partner's defection and never returns to cooperation.

The findings regarding proactive aggression must be interpreted with caution for several reasons. The sample tested in this study was a community sample and therefore sub-clinical. Proactive aggression appears more rarely than reactive aggression in the general population, and this was the case for our sample as well. The children in our sample generally scored very low on proactive aggression, with a maximum score of 6 out of the measure's maximum possible of 12, while 78 children out of the total 158 children for which proactive aggression scores were available had a score of zero. It is possible that our predictions for proactive aggression would appear in a clinical population. Also, interaction patterns might change with age and not continue into adolescence, although findings by Muñoz et al. (2008) who studied adolescents (ages 13-18) similarly show that reactive aggressive children are more responsive to low provocation. It is still possible however that the specific motives of the behavioral response are different in adolescents.

Finally, one hopeful take-away from this study is that aggressive behavior depends on the dyad. No matter what their levels of parent-rated aggression, children

began each interaction with a new partner at a similar level of cooperation and the order of the 80C and 80D partners did not matter. That is, no matter what children had experienced with a prior partner, they reset their expectations to begin interacting with the new partner. This is a somewhat surprising finding because one might predict that children high in reactive aggression would show anger or distrust towards everyone after one bad interaction. Instead both high proactive aggressive and high reactive aggressive start fresh with the new partner and modulate their play patterns based on the new interactions.

In summary, the current study makes the following key contributions. First, it validates the RPD as a rigorous, safe and child-friendly measure of behaviors in negative interactions. Combined, results from this study also suggest that high reactive aggressive children a) have a more prolonged reaction to partner defection b) fail to return to cooperation and engage in GRIM play even when the partner repeatedly attempts to restore cooperative interactions. The RPD shows promise as an experimental tool for further studying the motivations, self-protective or punitive, that produce this failure to restore cooperation. Finally, this study shows that aggressive behavior, reactive and proactive, is interaction dependent: children modulate their responses based on partner strategy and interaction characteristics and give new partners the benefit of the doubt.

CHAPTER 5

RESULTS: EXTENDING THE SIP MODEL TO POSITIVE INTERACTIONS

The SIP model has been primarily used for studying the cognitive steps involved in negative social interactions. The model however was proposed as a general theoretical framework meant to explain all types of interactions, including positive ones. This chapter presents findings on how the model applies to positive as well as negative social interactions. We then examine whether the responses to the SIP steps predicted behavior in the Repeated Prisoner's Dilemma (RPD).

According to the SIP model, how children think about particular social interactions drives the formulation of behavioral responses in a situation. Both the external stimulus (e.g., the provocation) and the response (e.g., the aggressive reaction) are represented in the mind with meaning (Dodge, 2008). The external stimulus acquires meaning through the intent that is attributed to the interaction partner: was their action accidental or on purpose, and in the case of negative outcome situations was their intention hostile or benign? Meaning is also conferred to the response by evaluating it in relation to, for example, feasibility or self-efficacy (am I able to carry it out?), desired consequences (will it be good for me, will it accomplish my goals?) and moral or social acceptability (is it socially and morally OK to behave like this?).

In order to evaluate the separate steps of this process of meaning making, SIP researchers typically present children with hypothetical vignettes and ask them to

imagine themselves in that situation. The vignettes describe a social interaction which results in either a positive or negative outcome but was caused by another person. For example, a child drops her homework on the way to school and another child steps on it in the mud. Given only this information, the action of the person who stepped on the homework is ambiguous – it may have been an accident or intentional. Children are then asked a series of questions to assess five steps in the SIP model (Crick & Dodge, 1994): 1) encoding situational cues (details about what happened), 2) attributing intent (whether the other person did it on purpose or not), 3) selecting goals (maintaining a friendly relationship or retaliation), 4) generating possible behavioral responses (socially competent or aggressive responses), 5) evaluating these responses (in terms of self-efficacy, consequences and moral acceptability) and selecting one to be enacted.

Most studies have focused on negative scenarios; the results repeatedly show correlations between the steps of the model and aggressive behavior. For example, both hostile intent attribution and positive evaluation of aggression have been connected to chronic aggressive behavior in real life (Bailey & Ostrov, 2008; Crick & Dodge, 1996; Dodge & Coie, 1987). A few studies have also linked prosocial behavior to socially competent responses on negative scenarios (Laible, McGinley, Carlo, Augustine, & Murphy, 2014; Laible, Murphy, & Augustine, 2014; Nelson & Crick, 1999). For example, Nelson and Crick (1999) found that prosocial 10 to 12 year-olds, identified as prosocial in peer nominations, were less likely to attribute hostile intent in provocation situations, were more likely to evaluate aggressive responses negatively and prosocial responses positively, and were more likely to adopt relational as opposed to instrumental

goals. In one rare study using positive outcome vignettes, Andrade et al. (2012) found that children with ADHD had impairments in processing social information for all the steps of the model. However, they have not drawn any connections between positive intent attribution and positive responses in the measure.

Given the dearth of studies using positive SIP vignettes, it remains unclear whether prosocial behaviors in positive interactions are formulated through the SIP steps. It is possible that children process positive interactions differently than negative ones. A new, large scale study of children's reciprocity finds that children are highly attuned to negative outcomes inflicted by another child and target the malefactor with punishment. By contrast, children seem to expect positive outcomes from others and do not strictly reciprocate positive interactions (Chernyak, Hu, Dunham & Blake, in prep). Children are also highly attuned to intentions, which affects their willingness to reciprocate with partners. This is evident in behavioral tasks with toddlers (Dunfield & Kuhlmeier, 2010) as well as in economic games with elementary school children (Sutter, 2007).

In the current study, we assessed children's SIP responses to both positive and negative vignettes. The positive vignettes were novel and negative vignettes were adapted from standard SIP tasks. The negative vignettes thus provided an anchor for validating children's responses, which could then be compared to the positive vignettes. We then explored possible links between SIP cognitions and strategies of play in the RPD game. Given that SIP studies rarely link children's thoughts to real behavioral measures, this last analysis was particularly valuable.

We had three main hypotheses which drove the analyses:

a) both the negative and positive outcome vignettes would generate responses consistent with the general SIP theory

b) hostile negative attribution would predict reactive aggression and positive evaluation of aggressive responses would predict proactive aggression

c) SIP responses would predict different levels of cooperation in the RPD.

Given that the relationship between the SIP and RPD has never been studied, we conducted several exploratory analyses. We first evaluated whether children processing social information more negatively (making more hostile attributions, generating more aggressive responses, and evaluating aggression more positively), as reflected by their total negative-vignettes score, are less likely to cooperate in the RPD. Conversely, we also tested whether children processing social information more positively (making more positive intent attributions, generating more positive responses and evaluating reciprocity more positively), as reflected by their total positive-vignettes score are more likely to cooperate in the RPD. Finally, we investigated whether positive and negative intent attribution, response generation or response evaluation would predict different levels of cooperation in the RPD.

Results

SIP responses to positive and negative vignettes

In our study children received the SIP measure first. They read three negative vignettes and three positive vignettes (for the full measure see Appendix D). Since we used the more traditional negative vignettes as an anchor to compare answers to positive

vignettes, children read the negative ones first. After each vignette, children answered a series of questions that focused on three of the SIP steps: intent attribution (step 2), response formulation (step 4) and response evaluation (step 5). Children responded to each question on a 4 point Likert scale. Scores were computed by averaging the relevant questions for each of the SIP steps tested.

On average, children attributed higher positive intent for ambiguous positive outcome interactions than negative intent for ambiguous negative outcome interactions (positive intent, positive outcome: $M = 2.178$, $SD = 0.558$; negative intent, negative outcome: $M = 1.223$, $SD = 0.735$, $t = 13.648$, $p < 0.001$). They also rated the likelihood of generating positive behavioral responses to a positive interaction as higher than the likelihood of generating negative behavioral responses (i.e., verbal or physical aggression) following a negative interaction (positive behavioral response: $M = 2.179$, $SD = 0.563$, negative behavioral responses: $M = 0.798$, $SD = 0.598$, $t = 20.297$, $p < 0.001$). Also, unsurprisingly, kids evaluated positive behaviors in response to positive interactions more positively than negative behaviors in response to negative interactions (evaluation of positive behavior: $M = 2.496$, $SD = 0.473$; evaluation of negative behavior: $M = 0.915$, $SD = 0.492$, $t = 27.151$, $p < 0.001$). This was true both on average and for each of the evaluation dimensions individually: self-efficacy, positivity of consequences and moral value of the behavior (see Table 13).

Table 13. Comparison of mean SIP responses for negative and positive outcome vignettes

	Negative		Positive		Paired t-tests		
	M	***SD***	***M***	***SD***	***t***	***df***	***p***
Intent attribution	1.222	0.735	2.178	0.558	13.64	163	<0.001
Behavioral response	0.798	0.598	2.179	0.563	20.29	163	<0.001
Evaluation	0.915	0.492	2.496	0.473	27.15	161	<0.001
Self-efficacy/Ease	1.263	0.705	2.367	0.545	14.69	163	<0.001
Consequences	0.865	0.556	1.528	0.483	27.85	164	<0.001
Moral acceptability/Value	0.624	0.542	2.601	0.508	31.01	163	<0.001

Negative SIP vignettes

To check whether the current SIP measure replicated findings from previous studies (e.g., see a meta-analysis by de Castro et al., 2002 and a cross-cultural study by Dodge et al., 2015), we used linear regression models to investigate the connection between hostile attribution bias (step 2) and the likelihood of formulating an aggressive response (step 4; see Table 14). In agreement with findings from previous studies, hostile attribution bias significantly predicted an increase in generating aggressive responses ($B = 0.298$, $p < 0.001$). We also checked whether positive evaluation of aggressive responses (step 5) was related to increased likelihood of generating such responses (step 4; Crick & Dodge, 1996; Dodge & Godwin, 2013; Dodge Laird, Lochman & Zelli, 2002; Dodge et al., 1997), and found that higher perceived self-efficacy in being aggressive, and higher perceived moral acceptability of aggression, but not positive evaluation of

aggression consequences significantly predicted higher likelihood of aggressive responses (self efficacy: $B = 0.251$, $p < 0.001$; moral acceptability: $B = 0.415$, $p < 0.001$).

Table 14. Predictors of selecting a aggressive response (step 4) in the negative SIP vignettes

	B	*SE*	95% CI	
Intercept	0.376	0.419		
Age	-0.050	0.040	-0.13	0.03
Gender	0.013	0.064	-0.11	0.14
Negative intent	0.299	0.046***	0.21	0.39
Self-efficacy	0.251	0.053***	0.15	0.35
Consequences	-0.036	0.080	-0.20	0.12
Moral acceptability	0.415	0.088***	0.24	0.59
# Children	163			
R2	0.56			
F(6,156)	33.44			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Positive SIP vignettes

To further investigate whether processing of social information functions in analogous ways when interactions and their outcomes are positive, we used linear regression models to mirror the analyses we conducted for responses to the negative vignettes (see Table 15). We found that increased attribution of positive intent significantly predicted increased likelihood of generating a positive response ($B = 0.351$, $p < 0.001$). Also a higher perceived ease of responding positively was significantly related to higher likelihood of generating a positive response ($B = 0.433$, $p < 0.001$). Neither positive evaluation of the consequences nor their perceived moral acceptability

(which were unsurprisingly very high) influenced significantly the likelihood of generating a positive response.

Table 15. Predictors of selecting a positive response (step 4) in the SIP positive vignettes

	B	*SE*	95% CI	
Intercept	1.281	0.461		
Age	-0.090	0.044	-0.18	0.00
Gender	0.057	0.072	-0.08	0.20
Positive intent	0.351	0.068***	0.22	0.48
Self-efficacy	0.433	0.110***	0.21	0.65
Consequences	-0.159	0.142	-0.44	0.12
Moral acceptability	0.150	0.130	-0.11	0.41
# Children	163			
R2	0.37			
F(6, 156)	15.26			

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Age and social information processing

We then investigated the relation between children's age and gender and SIP both in negative and positive social interactions. Age significantly predicted attribution of negative intent in negative interactions, with older children making fewer negative intent attributions ($B = -0.166$, $p = 0.015$). Attribution of positive intent in positive interactions was not significantly predicted by age ($B = 0.098$, $p = 0.063$). Neither the generation of negative behavioral responses (i.e., verbal or physical aggression), nor positive behavioral responses differed with age (negative behavioral response: $B = -0.062$, $p = 0.271$; positive behavioral response: $B = -0.026$, $p = 0.630$). Older children evaluated

both the consequences and the moral acceptability of aggressive responses in negative interactions more positively (consequences: $B = 0.122$, $p = 0.019$; moral acceptability: $B = 0.121$, $p = 0.018$). There were no age differences in the evaluation of positive responses to positive interactions (self-efficacy/ease: $B = 0.063$, $p = 0.221$; consequences: $B = 0.036$, $p = 0.423$; moral value: $B = 0.083$, $p = 0.080$). Gender was not significantly related to any of the SIP steps for neither negative nor positive social interactions.

In sum, the positive and negative SIP vignettes produced results that were consistent with prior results. Notably, children children appeared to use similar processing steps in the positive and negative vignettes, with intent attribution and self-efficacy predicting response selection in both cases.

Social information processing and reactive and proactive aggression

Previous studies have found a relationship between hostile attribution bias (negative intent) and reactive aggression, and positive evaluation of aggressive responses and proactive aggression. To replicate these studies we used regression models to check the relation between different SIP steps in negative interactions and the PRPA measure (continuous scores). We found that the likelihood of generating aggressive responses as measured by the SIP measure significantly predicted the overall aggression (reactive and proactive combined) measured through the PRPA ($B = 1.041$, $p = 0.020$). However, we failed to replicate the differential associations between specific SIP steps and reactive and proactive aggression. Reactive aggression was not significantly predicted by negative

intent attribution (B = -0.129, p = 0.640), and proactive aggression was not significantly predicted by positive evaluation of aggression (B = 0.634, p = 0.207).

Positive and Negative SIP responses and the RPD

Finally, we investigated the relation between SIP measured through the negative and positive vignettes and cooperative behavior in the Prisoner's Dilemma game. We first computed an overall SIP score (range: 0 to 3) for the negative vignettes indicating the degree to which each child processed social information negatively by making negative intent attributions, generating aggressive responses and evaluating aggression positively. We used a regression model with children's decisions in the RPD as a dependent variable and the overall SIP score for the negative vignettes as an independent variable. Age and gender were introduced as covariates. We found no relationship between the negativity of SIP for negative vignettes and the likelihood of cooperating (or defecting) in the RPD game (B = -0.016, p = 0.179). We also computed an analogous overall SIP score for the positive vignettes (range: 0 to 3), indicating the degree to which each child processed social information positively by making positive intent attributions, generating positive behavioral responses and evaluating these responses positively. We used a similar regression model with children's decisions in the RPD as the dependent variable and the overall SIP score for positive vignettes as the independent variable, controlling for age and gender. Again, we found no relation between this overall positivity score and cooperation in the RPD game (B = 0.327, p = 0.208). Lastly, we checked whether the cooperation in the PD game was predicted by any of the measured

SIP steps for the negative vignettes and then separately for the positive vignettes. We again found no significant relationships between any of the variables (see Table 16).

Table 16. Social information processing steps and cooperation in the RPD

	Negative vignettes				Positive vignettes			
	B	*SE*	95% CI		*B*	*SE*	95% CI	
Intercept	-4.088	1.018***			-4.879	1.064***		
Age	0.339	0.101**	0.14	0.54	0.345	0.100**	0.15	0.54
Gender	0.099	0.166	-0.23	0.43	0.097	0.168	-0.23	0.42
Intent attribution	0.032	0.135	-0.23	0.30	0.077	0.166	-0.25	0.40
Behavioral response	-0.148	0.202	-0.55	0.25	0.069	0.190	-0.30	0.44
Evaluation	0.150	0.222	-0.28	0.58	0.181	0.218	-0.25	0.61
Log Likelihood	-3019.142				-3124.211			
# Observations	4890				4890			
# Children/Clusters	163				163			
$\chi 2$(df)	14.03(5)				16.37			
Pseudo R2	0.01				0.02			

Discussion

The findings of this study partially supported the idea that the SIP model can also explain behavior generation in positive social interactions. Results confirmed our first hypothesis, that the new positive outcome vignettes and updated negative outcome vignettes both generated responses consistent with SIP theory. However, other prior results failed to replicate. Namely, we found no relationship between reactive and proactive aggression and a hostile attribution bias or the overall SIP scores. Additionally, we found no indication that the SIP results generalized to situations outside the vignette

scenarios. Responses to both the positive and negative vignettes did not predict different levels of cooperation in the RPD. However, our analyses did reveal intriguing effects of age on SIP cognitions.

Our findings verified that the SIP questionnaire developed for this study was a valid measure of the SIP framework. We replicated previous results for negative vignettes in the literature, finding that attribution of negative/hostile intent significantly predicted aggressive responses, as well as more positive evaluation of aggression (Crick & Dodge, 1996). The positive vignettes were novel and also showed response patterns consistent with SIP theory. Combined these results provide a validation that the vignettes used here tested the same cognitive constructs as previously used questionnaires.

This study's main contribution to the SIP literature is testing the model's mechanisms for positive outcome interactions. This was accomplished by using an extended SIP questionnaire that includes positive-outcome vignettes. We showed that positive intent attribution is related to an increase in generating positive responses. Also, we showed that the perceived ease in carrying out positive responses was associated with a higher likelihood of generating such responses.

The relationship between positive intentions and reciprocity has been studied in developmental psychology (e.g., Vogelsang & Tomasello, 2016; Vaish, Hepach, & Tomasello, 2017; Warneken & Tomasello, 2013). Children distinguish between different types of intentions and base their decisions to reciprocate not on outcome alone but on the underlying intention as well (Dunfield & Kuhlmeier, 2010). Although the SIP results focus on hypothetical scenarios, they show that even in ambiguous situations in which

intentions are not clear or directly observable, positive reciprocity still depends on how the situation is represented in the mind: as having positive or neutral intent. This is notable given that previous studies have found that attribution of intent is reduced when the outcome is positive (Leslie, Knobe, & Cohen, 2006). These results are consistent with the SIP model.

Unsurprisingly, children had high positive evaluations (almost at ceiling) of positive responses both in terms of consequences and moral value. This had no effect on their positive response selection within the SIP model. Perceived ease in carrying out the positive response was also related to positive responding. However, it is unclear how children interpreted this question: "would it be easy or hard for you to say or do something nice to the other child?". Did they think of their ability to be nice or how costly it would be for them to be nice to the other child? We tried to keep our questions consistent across positive and negative scenarios, but our results suggest that for the generation of positive behaviors in positive interactions, other evaluation criteria might be relevant than for negative behaviors. Even though response evaluation might be an essential component of all response formulation, the relevant evaluation criteria might be more situation-specific. Future SIP measures should be developed to investigate which evaluation criteria are relevant for positive response selection.

The results from the negative outcome SIP vignettes allowed us to test the relationships between hostile attribution and reactive aggression, and between positive evaluations of aggression and proactive aggression that have been found in other studies (Bailey & Ostrov, 2008; Crick & Dodge, 1996; Kempes et al., 2006). We found that

children's aggressive response selection in the vignettes significantly predicted parent report of overall aggression (PRPA; Kempes et al., 2006). However, we found none of the predicted relationships for reactive and proactive aggression specifically. Intriguingly, other studies have also failed to replicate these results. For example, Dodge et al. (1997) found no differences in hostile attribution between reactive and proactive aggression. Kempes et al. (2006) also did not find correlations between positive evaluations of aggression consequences and proactive aggression. We consider three possible explanations for these replication failures.

First, methodological reasons could be responsible for this null result. Studies of SIP and reactive and proactive aggression have relied on a variety of measures (for a review of reactive and proactive aggression measures, see Polman et al., 2007). Reactive and proactive aggression measures have taken different forms, from questionnaires to observation-based measures. They have also relied on different informants (teachers, parents, self), with cross-informant agreement usually being very low. Also, they have sometimes distinguished not just between the function of aggression (reactive or proactive) but also the form (physical or relational). According to Polman et al., the distinction between physical and relational aggression may be a more relevant distinction than reactive and proactive aggression.

Second, SIP measures have also varied a lot in terms of their form (text-based, video-based, real interaction), the number of SIP steps they test, whether the same vignette is used to test multiple steps or just one, the hypothetical scenarios presented (provocation situations or peer rejection situations), the types of questions asked (open

ended or multiple choice), and the way the questions are formulated. A review and meta-analysis by de Castro et al. (2002) highlights further differences in SIP measures of intent attribution. Given the wide variety of measures used, the absence of conclusive results could be due to methodological diversity, although it is difficult to pinpoint the relevant feature responsible. For example, it is possible that children require richer interactions involving perhaps facial cues, or deeper knowledge of the other child in order to make relevant intent attributions. A systematic review would be needed to clarify which methodological features play an important role in the way SIP measurements and reactive-proactive aggression measurements are connected.

Lastly, it is also possible that differences in the population tested might account for our failure to replicate the findings. Our participants represent a non-clinical sample with relatively low levels of reactive and proactive aggression. Although this is a plausible explanation, we believe it is unlikely given that prior work showing the relationship between reactive and proactive aggression and SIP processing also used a community sample (Crick & Dodge, 1996). It is also important to note that children from different SES backgrounds or children of different ethnicities might produce different kinds of interpretations of the same social interaction, thus further research is needed to understand how other factors might mediate the link between SIP and aggression.

We had hypothesized that a predisposition to attribute positive intent in ambiguous positive interactions in the SIP may predict cooperation in the RPD. However, none of the SIP steps for either negative or positive scenarios were related to decision patterns in the RPD. Given that children's perceptions of their partner's motivations

influenced their cooperation in the RPD (see Chapter 3), it seemed plausible that children's positive intent attributions in SIP measures would also be related to cooperation in the RPD. However, this relationship did not emerge across the tasks and differences between the SIP vignettes and the RPD may help to explain the null result. In the RPD game, the outcome always depends on the combined actions of the two partners, whereas in the SIP vignettes the interaction partner is the only one responsible for the action. In the RPD game, cooperation can be self-interested, whereas in the SIP vignettes there is no direct incentive for responding positively. In the RPD game, the interaction is repeated while the SIP vignettes describe a single stimulus and a single response. Thus it is possible that cooperation in the RPD depends on a more intricate set of cognitions than simple intent attribution, which motivates positive responses in the SIP scenarios.

Our analyses did yield some intriguing results regarding age differences in SIP responding. We found that hostile intent attribution decreased with age, but found no age differences in positive intent attributions. This is in agreement with the proposal that inferring benign intent is more developmentally advanced and cognitively taxing than inferring hostility (Dodge, 2006). Better cognitive capacities, learning and socialization are all needed to override the more primitive representation that negative personal outcomes always happen through another's hostile intentional action. Cognitive development related to the contemplation of alternatives and thinking about hypothetical possibilities could be responsible for different patterns of attributional biases (Dodge, 2006), and might explain the decrease in hostile attribution bias between 9 and 11 years of age.

The evaluation of consequences and moral acceptability of aggression were also related to age, with older children giving more positive evaluations of aggression. This is in accordance with the results of Huesmann and Guerra (1997) who, in a large scale longitudinal study found that elementary school children tended to approve more of aggression with age, which was also related to an increase in their aggressive behavior. Note however, that these findings were obtained from measures other than vignette-based SIP questionnaires and children in this study came from low SES, high-crime neighborhoods, so it is unclear whether the same factors contribute to an increase with age in positive evaluation of aggression in our sample. Unlike this study we found no changes in parent-reported aggression with age.

In summary, we found partial evidence for the SIP model's consistency in explaining how children think about both positive and negative social interactions. We showed that positive outcome vignettes are generally eliciting responses consistent with the SIP framework, even though more work is needed to further specify the last step of the SIP model: how positive responses are evaluated. However, this study found no connections between the SIP measure and proactive and reactive aggression or cooperative behavior in the RPD. This suggests that further research is needed for establishing whether the SIP model is generalizable to situations outside the hypothetical scenarios task.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS: TRAITS, COGNITIONS, AND BEHAVIORS IN SOCIAL INTERACTIONS

This dissertation has shown how traits, cognitions and behaviors that are relevant for social interactions can be studied in an integrative way by combining approaches and measures from developmental psychology, behavioral economics and game theory. We used aggression and its varying manifestations in social interactions as a point of focus. With the goal of studying live negative social interactions in a safe and rigorously quantifiable way, we first explored how the RPD, a measure from behavioral economics and game theory, could be used as a minimal social interaction. We showed that decisions in the game are motivated by rich social cognitions and that patterns of play reveal strategic behavior that can be analyzed in terms of game theoretic principles. We then showed that patterns of decisions in the RPD can also capture behavioral signatures of different aggression types, reactive and proactive. More specifically, we found that reactive aggression is characterized by lack of return to cooperation following partner defection, indicating persistent and unforgiving reactivity to negative social experiences. Finally, we researched the SIP theoretical model, which has successfully described how various cognitions lead to the formulation of aggressive behaviors in social interactions, and showed how it might be relevant for explaining the formulation of positive social behaviors as well. Together these findings constitute a promising start for the unification

of cognitive and behavioral measures from two different research traditions and the unification of theoretical models of positive and negative social interactions.

For this study we developed a novel, computerized version of the RPD based on a visual interface that has been used successfully with children (Blake et al., 2015). In the standard PD, two players must simultaneously decide to either Cooperate (C) or Defect (D). The combined decision determines how many points each player gets. In the current version, mutual cooperation (CC) pays 3 points each and mutual defection (DD) pays 1 point each. However, if one player cooperates and the other defects (CD or DC), the cooperator receives zero and the defector receives 4 points. Over multiple rounds, mutual cooperation can lead to high payoffs, but there is always a short-term incentive to defect.

Chapter 3 described how children construe this RPD task. Game theory has made various predictions regarding the behavior of rational agents in the RPD, based on game characteristics: the number of rounds, the payoff structure *etc*. Studies with adults have observed behaviors in the RPD and cognitions have been inferred from experimental manipulations of the game set-up. However, only a relatively small number of studies have directly probed cognitions and motivations in the RPD. No such studies had been previously done with children.

In order to use the RPD to study maladaptive social behaviors such as aggression, we first needed to know if children thought of the RPD as a real social interaction. This dissertation shows that when asked to motivate their decisions in the game, children made reference to very rich social cognitions focused on trust, reciprocity, joint gain, personal interest, taking advantage of the partner, prosocial reasons, and personality traits

(belonging to themselves or the partners). This verifies that children interpret the RPD as a real social interaction. Their motivations were also related to their behavior in the game. Children who made reference to interpersonal motivations, explaining their decisions in the game in relation to those of the partner, were more likely to cooperate. Also perceiving the partner's motivations as being more interpersonal lead to increased cooperation.

Probing children's interpretations of their decisions in the RPD is crucial for informing our own interpretation of their behavior. For example, defection in the game could mean either "I'm playing it safe" or "I'm exploiting my partner". Repeated defection could mean "I don't trust you" or "I'm angry with you". More work is needed to gain a deep understanding of how children view specific situations in the game. In this study we were only able to ask children why they decided to cooperate or defect in general. (We asked in a neutral language about pushing and pulling, the two actions in the game that amounted to cooperation and defection respectively). However, cooperation and defection in a repeated game are modulated by the specific history of that interaction, and in principle they could mean different things at different points in the game. It is possible that the child defects at first to play it safe but after they have been reassured that the partner is cooperative they realize they can take advantage of the partner's kindness. Future research is needed to get that level of resolution for our understanding of how the RPD interaction is interpreted by children.

We have also shown that children approach the RPD in a strategic fashion. In Chapters 3 and 4, we show how children's sequences of play approximate, above chance

levels, standard strategies from game theory. Blake et al. (2015) have previously shown that children play contingently: their cooperation in the game is dependent on the partner's actions. We replicated those findings through analyses described in Chapter 4. Moreover, we were able to show that patterns of play in the game can be related to types of aggressive behavior in real life. This is important not just for validating the RPD as a measure of social interactions, but also for creating novel ways to rigorously quantify social behavior in the context of child psychopathology.

Maladaptive social interactions both characterize and represent a risk for psychopathology in children, and one of the most common ways in which social interactions go wrong in childhood involves aggressive behavior. In chapter 4 we showed how children rated high on reactive aggression by their parents defected in response to partner defection, and failed to return to cooperation even after repeated attempts from a generally cooperative partner to restore mutual cooperation. High reactive aggressive children preferred the unforgiving game theoretic strategy GRIM, in which the player never returns to cooperation following partner defection. These findings suggest that reactive aggressive children not only engage in negative interpretation of–and negative response to–certain social behaviors, but that they also have difficulty in moving past negative interactions and resolving conflict. This could be because of a propensity towards persistent anger, or because of difficulty with restoring trust in a partner that has displayed negative action. Both of these options find support in the previous literature (Sharp et al., 2011; Waschbusch et al., 2002). However, our results add to this literature by showing that even after a GRIM response to the partner's defection, high reactive

children are able to start fresh with a new partner. Therefore, if anger is the underlying cause of an aggressive reaction, it is highly targeted toward individual actors. Further research is needed to elucidate this matter, and experimental manipulations with the RPD task would be a fruitful means fur further inquiry.

The ability to capture signatures of aggression in RPD play patterns is also significant for longer-term research goals involving more generally the study of social interactions in the context of child psychopathology. Traditional measures of maladaptive social behavior are primarily report measures documenting either past experiences or reactions to hypothetical scenarios and do not measure real behavior as it happens. They rely heavily on language abilities and are difficult to standardize across ages. The RPD on the other hand measures behavior as it happens in real, incentivized interactions, can be used across a wide age range and is a practical tool for functional brain imaging and ERP studies (e.g., Rilling et al., 2007). The RPD also offers advantages over the other few live interaction tasks that have been used for studying aggression: it is not a purely competitive task, it allows for both positive and negative interactions to occur, and it provides an ongoing signal of the partner's simultaneous decisions. The RPD is a versatile measure that can be used to study a variety of behaviors and cognitions, both positive and negative. Thus the RPD could also be used to standardize the measurement of negative and positive social behaviors in development.

More broadly, this dissertation describes an attempt to breach the divide between the study of positive and negative social interactions. We have only been partially successful in accomplishing this goal, as we have shown that responses to SIP measures

with positive-outcome vignettes are consistent with SIP theory, but failed to find statistically significant connections between these responses and real, cooperative behavior in the RPD.

The SIP model has been used to explain the thought processes involved in formulating aggressive behaviors, but the framework was intended as a more general explanation for how *all* social behavior is produced. The framework proposes several information processing steps: encoding of situational cues, attribution of intent, goal selection, response generation, response evaluation and finally response selection and production. The model has been generally tested through hypothetical vignettes describing a social interaction, followed by questions meant to capture cognitions that appear during every step of the response formulation process: how many social cues are encoded? Is the action in the vignette perceived as hostile or benign? What goal is adopted for the situation, relational or instrumental? Finally, what are the possible behavioral responses that come to mind, and how positively are they evaluated? Attribution of hostile intent and positive evaluation of aggression in these hypothetical scenarios have been linked to real life aggressive behavior.

In this study we wanted to also test whether the same steps are involved in the formulation of positive behaviors in positive social interactions, and how they might relate to prosocial behaviors in the RPD. To accomplish this, we created an extended SIP measure that included not only the traditional negative-outcome vignettes but also novel, positive outcome-vignettes in which the action of the partner was ambiguous. This allowed us to measure whether the positive action was perceived as purposeful or

accidental, whether the action was reciprocated or not and how positively was reciprocity evaluated.

In the formulation of positive responses to these hypothetical scenarios, we found mechanisms consistent with SIP theory: attribution of positive intent significantly predicted social reciprocity. Positive evaluation in terms of consequences and moral value of the response were at ceiling, but the perceived ease in carrying out the positive response significantly predicted positive reciprocity. However, these responses for the SIP vignettes failed to predict cooperative behavioral patterns in the RPD. We also failed to replicate previous findings that linked reactive and proactive aggression to hostile attribution and positive evaluation of aggression in the negative scenarios. Further work is needed to establish why no connections were found. However, it remains plausible and likely that children's thoughts are linked to their behavior only within the specific tasks used. Therefore, testing children perceptions of their partner's intentions in the RPD may be a better predictor of their decisions in that task.

In summary, the study of social interactions, which has been the topic of this dissertation, is critical for ensuring children's well-being and life success. It is therefore important to quantify maladaptive social cognitions and behaviors in a rigorous parametric manner. We have shown that strategic games, such as the RPD are exemplary tools for doing this. The continued study of social interactions in strategic games can thus provide a foundation for the development of targeted interventions and therapies for pathological social behaviors. Further, while negative interactions have received more attention due to their health impact, it is also important to study positive social

interactions, their cognitive underpinnings, and how positive social behaviors can be encouraged and maintained. Finally, to obtain a unified view of both negative and positive interactions, the merging of approaches from different research traditions is not only fruitful, as we have shown in this dissertation, but may also be necessary.

APPENDIX A

Instructions and comprehension checks for the RPD measure

An image of the game board (see Fig. 2) appears on each screen that becomes partially interactive when the explanation requires it. Children receive the following instructions that self-guide them through a demonstration of the task and check their comprehension. Children must pass the comprehension checks before proceeding to the testing phase.

This is what you will see on your screen. The hand below is you. The hand at the top is your partner. Click on the button below to move to the next screen.

You can make your hand push the trays or pull them.

Click the "Push" button below to see what happens.

The coins that end up on the green area at the bottom go to you. The coins that end up on the green area at the top go to your partner. The coins that get stuck on the red area belong to no one and they disappear.

When you click the Push button, your partner gets 3 coins and 1 coin gets stuck on the red area and no one gets it. Click on the button below to move to the next screen.

Now click the "Pull" button to see what happens.

When you click Pull, you get 1 coin and 3 coins get stuck on the red area and no one gets those. Click on the button below when you are ready to move to the next screen.

Your partner can also make their hand push the trays or pull them. Click on the “View” button below to see what happens. If your partner clicks Pull, then they get 1 coin. If they click Push, then you get 3 coins. Click on the button below when you are ready for some practice rounds.

These are practice rounds and they are not part of the real game. This is just to show you what can happen in the game. In the real game you will be playing 10 rounds with each partner.

Please start by clicking on the "Push" button.
Great! You clicked on the “Push” button and your partner clicked on the “Push” button. You got 3 points and your partner got 3 points.

Click on the “Push” button again.
Great! You clicked on the “Push” button and your partner clicked on the “Pull” button. You got 0 points and your partner got 4 points.

Now click on the “Pull” button.
Great! You clicked on the “Pull” button and your partner clicked on the “Push” button. You got 4 points and your partner got 0 points.

Click on the “Pull” button again.

Great! You clicked on the “Pull” button and your partner clicked on the “Pull” button. You got 1 point and your partner got 1 point.

Comprehension check:

The child needs to pass the comprehension checks before being able to play the game. The platform will not allow the child to proceed unless they get the questions right. Children are prompted to type their answer into text boxes for their points and the partner’s points.

If they child inputs the wrong answer the following message appears:

Incorrect! Please try again...

If the child inputs the correct answer the following message appears:

Correct!

If you click “Push” and your partner clicks “Push” how many do you get? And how many does your partner get?

If you click “Push” and your partner clicks “Pull” how many do you get? And how many does your partner get?

If you click “Pull” and your partner clicks “Push” how many do you get? And how many does your partner get?

If you click “Pull” and your partner clicks “Pull” how many do you get? And how many does your partner get?

What combination of buttons delivers the most points for you and your partner combined?

□ You: Push Partner: Push

□ You: Push Partner: Pull

□ You: Pull Partner: Push

□ You: Pull Partner: Pull

How many rounds are you going to play with each partner?

□ 2

□ 3

□ 8

□ 10

Once you are ready to begin the game press the button below and we will connect you to another child over the internet.

APPENDIX B

Summary of coding scheme for motivations in the RPD

After the RPD task, children were asked to type open-ended responses to the following questions:

1. We are interested in how you played the game. Please describe why you clicked the push and the pull buttons.

2. Sometimes your partner pressed "pull". Why do you think they did that?

3. Sometimes your partner pressed "push". Why do you think they did that?

A coding scheme was developed by the author and thesis supervisor based on motivations and reasons for decisions that have been examined for adults who have done similar cooperative dilemma tasks (Dreber et al., 2014). These a priori coding categories were then checked against the responses to determine whether categories needed to be added, modified or removed. As a result of this process two categories were combined: competitive defection, safety-motivated defection since the wording of children's answers did not allow for us to make a clear distinction between the two. The final categories were intended to capture all unique motivations that would affect decisions in the RPD and thus children could mention multiple motivations for each question.

Children's answers to the questions were coded by two independent coders who were blind to any other information about the child including, their age, gender and decisions in the game. Answers were assigned a code of 1 for each type or subtype of motivation the child mentioned and a code of 0 if the type was not mentioned. Children who mentioned more that one motivation received a code of 1 for each type or subtype

they mentioned. Average Cohen's kappa for inter-coder reliability was 0.72.

Disagreements were resolved by discussion between the two coders.

Interpersonal motivations	*At least one of the 4 subcategories below was mentioned*
Trust	Decisions are motivated by trust in the partner or are an attempt to test whether the partner is trustworthy and will cooperate
Reciprocity	Decisions are an attempt for players to mirror each other, either by mirroring the partner's response or trying to elicit a reciprocal response from the partner
Fairness	Decisions are a way to ensure fairness, equality or to share points
Joint gain	Decisions are motivated by mutual benefits from cooperation

Strategic self interest	*At least one of the 2 subcategories below was mentioned*
Personal gain	Decisions are made to get points, to win or to prevent the partner from winning
Advantage taking	Decisions are meant to trick the partner and take advantage of their generosity or gullibility

Prosocial	*Decisions are motivated by positive personality or character traits: being nice, kind, helpful, generous etc.*

APPENDIX C

The Parent-rating scale of Reactive and Proactive Aggression (PRPA)

The measure was developed by Kempes, Matthys, Maassen, van Goozen, & van Engeland (2006). The procedures for creating and validating the measure can be found in this paper and the measure can be retrieved from the paper's appendix. The questionnaire consists of 11 items, 6 items measuring reactive aggression and 5 items measuring proactive aggression. Parents are asked to indicate the frequency of different types of behaviors on a 3-point Likert scale, choosing between: never (0), sometimes (1) and often (2).

Table C1. Descriptive statistics for the PRPA

	Mean	**SD**	**Median**	***N* of kids with median score**	**Actual Range**	**Min-Max**
Reactive aggression	2.95	2.93	3	19	0-10	0-12
Proactive aggression	0.92	1.26	1	47	0-6	0-10

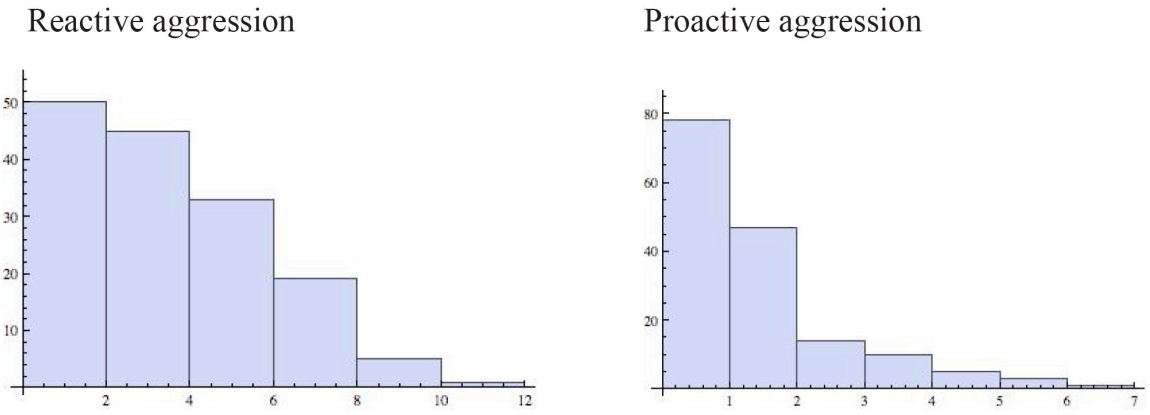
We compared scores from the sample described in this dissertation with those reported by Kempes, de Vries, Matthys, van Engeland, & van Hooff (2008). Our findings are consistent with those found by Kempes et al. for the non-clinical population (normal controls) in their study (see Table C2).

Table C2. Comparisons between current data and Kempes et al. (2008). Scores were transformed to fall within a range of 0 to 1. T-tests show differences between the current data (used in this dissertation) and results by Kempes et al.

	Current data		**Kempes et al. 2008**					
			Normal controls			Disruptive behavior disorder		
	N	Mean (SD)	*N*	Mean (SD)	*t*	*N*	Mean (SD)	*t*
Reactive aggression	153	0.25 (0.20)	39	0.31 (0.23)	1.62	39	0.68 (0.24)	11.49***
Proactive aggression	158	0.09 (0.13)	39	0.07 (0.09)	0.91	39	0.33 (0.21)	9.00***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
scores transformed to fall into a 0-1 range

Figure C1. Distribution of scores for the PRPA measure



APPENDIX D

SIP QUESTIONNAIRE

Negative-outcome vignettes:

1) Imagine that you are on your way to school one morning. As you are walking a homework paper falls out of a notebook that you are carrying. A kid you know walks by and steps on the paper, leaving a muddy footprint right across the middle. The kid looks at your homework paper then looks at you.

2) Imagine it's a hot day and you just got an ice cream cone. You are holding your ice cream in your hand and are about to eat it. A kid you know bumps into you and your ice cream falls on the pavement. The kid looks at the melting ice cream then looks at you.

3) Imagine you are in class one day. The teacher asks you to go write something on the board. A kid you know has his foot in the aisle and you trip over it. The kid looks at his foot then looks at you.

Positive-outcome vignettes:

1) Imagine you are on the playground one afternoon. You are walking past a group of kids who are having a lot of fun playing with a soccer ball. A kid you know kicks the ball in your direction and the ball lands near you. The kid looks at the ball then looks at you.

2) Imagine that you are in art class one day. You are painting a landscape and you need blue paint for the sky, but you can't find the container. A kid you know walks by

carrying a jar of blue paint and puts it down next to you. The kid looks at the blue paint then looks at you.

3) Imagine that you are in the cafeteria one day. You are holding your lunch tray and looking for a seat but the cafeteria is full and all seats are taken. A kid you know stands up and clears the table in front of him. The kid looks at the empty seat then looks at you.

Questions:

Intent attribution:

	not at all	a little	much	very much
1. How much do you think the kid did this on purpose?	□	□	□	□
2. How much do you think the kid was trying to be mean/nice?	□	□	□	□

Response generation:

	definitely would not	probably would not	probably would	definitely would
3. Would you do or say something mean/nice to the kid in the future?	□	□	□	□
4. Would you do or say something mean/nice to the kid right away?	□	□	□	□

Response evaluation:

a) Self-efficacy/Ease:

	very easy	pretty easy	pretty hard	very hard
5. Would it be easy or hard for you to do or say something mean/nice to the kid in the future?	□	□	□	□
6. Would it be easy or hard for you to do or say something mean/nice to the kid right away?	□	□	□	□

b) Consequences:

	very well	pretty well	pretty bad	very bad
7. Would things turn out well or badly if you did or said something mean/nice to the kid in the future?	□	□	□	□
8. Would things turn out well or badly if you did or said something mean/nice to the kid right away?	□	□	□	□

c) Moral acceptability/value:

	very good	pretty good	pretty bad	very bad
9. Do you think it is good or bad to do or say something mean/nice to the kid in the future?	□	□	□	□
10. Do you think it is good or bad to do or say something mean/nice to the kid right away?	□	□	□	□

Table D1. Internal consistency of the measure across the different vignettes for each of the steps of the SIP model

	Cronbach's α	
	Negative vignettes	Positive vignettes
Intent attribution	0.85	0.69
Response generation	0.82	0.76
Response evaluation		
Self-efficacy/Ease	0.84	0.80
Consequences	0.86	0.86
Moral acceptability/value	0.85	0.85

REFERENCES

Andrade, B. F., Waschbusch, D. A., Doucet, A., King, S., MacKinnon, M., McGrath, P. J., ... & Corkum, P. (2012). Social information processing of positive and negative hypothetical events in children with ADHD and conduct problems and controls. *Journal of Attention Disorders*, *16*(6), 491-504.

Andreoni, J. & Miller, J. H. Rational cooperation in the infinitely repeated prisoner's dilemma: Experimental evidence. *Econ. J.* **103,** 570–585 (1993).

Atkins, M. S., Osborne, M. L., Bennett, D. S., Hess, L. E., & Halperin, J. M. (2001). Children's competitive peer aggression during reward and punishment. *Aggressive Behavior*, *27*(1), 1-13.

Atkins, M. S., Stoff, D. M., Osborne, M. L., & Brown, K. (1993). Distinguishing instrumental and hostile aggression: does it make a difference?. *Journal of abnormal child psychology*, *21*(4), 355-365.

Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science*, *242*(4884), 1385-1390.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390-1396.

Bailey, C. A., & Ostrov, J. M. (2008). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence*, *37*(6), 713-722.

Bell, D. J., Luebbe, A. M., Swenson, L. P., & Allwood, M. A. (2009). The children's evaluation of everyday social encounters questionnaire: Comprehensive assessment of children's social information processing and its relation to internalizing problems. *Journal of Clinical Child & Adolescent Psychology*, *38*(5), 705-720.

Blake, P. R., Rand, D. G., Tingley, D., & Warneken, F. (2015). The shadow of the future promotes cooperation in a repeated prisoner's dilemma for children. *Scientific reports*, *5*.

Camodeca, M., & Goossens, F. A. (2005). Aggression, social cognitions, anger and sadness in bullies and victims. *Journal of Child Psychology and Psychiatry*, *46*(2), 186-197.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. G. (2000). Prosocial foundations of children's academic achievement. *Psychological science*, *11*(4), 302-306.

Card, N. A., & Little, T. D. (2006). Proactive and reactive aggression in childhood and adolescence: A meta-analysis of differential relations with psychosocial adjustment. *International Journal of Behavioral Development*, *30*(5), 466-480.

Catalano, R. F., Berglund, M. L., Ryan, J. A., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The annals of the American academy of political and social science*, *591*(1), 98-124.

Clark, C. B., Thorne, C. B., Hardy, S., & Cropsey, K. L. (2013). Cooperation and depressive symptoms. *Journal of affective disorders*, *150*(3), 1184-1187.

Connor, D. F., Steingard, R. J., Cunningham, J. A., Anderson, J. J., & Melloni Jr, R. H. (2004). Proactive and reactive aggression in referred children and adolescents. *American Journal of Orthopsychiatry*, *74*(2), 129.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological bulletin*, *115*(1), 74.

Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child development*, *67*(3), 993-1002.

Dal Bó, P. (2005). Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *American economic review*, *95*(5), 1591-1604.

Dal Bó, P., & Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, *101*(1), 411-29.

de Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child development*, 916-934.

Dodge, K. A. (2006). Translational science in action: Hostile attributional style and the development of aggressive behavior problems. *Development and psychopathology*, *18*(3), 791-814.

Dodge, K. A. (2008). On the meaning of meaning when being mean: commentary on Berkowitz's "On the Consideration of Automatic as Well as Controlled Psychological Processes in Aggression". *Aggressive behavior*, *34*(2), 133-135.

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of personality and social psychology*, *53*(6), 1146.

Dodge, K. A., & Newman, J. P. (1981). Biased decision-making processes in aggressive boys. *Journal of Abnormal Psychology*, *90*(4), 375.

Dodge, K. A., Godwin, J., & Conduct Problems Prevention Research Group. (2013). Social-information-processing patterns mediate the impact of preventive intervention on adolescent antisocial behavior. *Psychological Science*, *24*(4), 456-465.

Dodge, K. A., Laird, R., Lochman, J. E., & Zelli, A. (2002). Multidimensional latent-construct analysis of children's social information processing patterns: Correlations with aggressive behavior problems. *Psychological assessment*, *14*(1), 60.

Dodge, K. A., Lochman, J. E., Harnish, J. D., Bates, J. E., & Pettit, G. S. (1997). Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth. *Journal of Abnormal Psychology*, *106*(1), 37.

Dodge, K. A., Malone, P. S., Lansford, J. E., Sorbring, E., Skinner, A. T., Tapanya, S., ... & Bacchini, D. (2015). Hostile attributional bias and aggressive behavior in global context. *Proceedings of the National Academy of Sciences*, *112*(30), 9310-9315.

Dreber, A., Fudenberg, D., & Rand, D. G. (2014). Who cooperates in repeated games: The role of altruism, inequity aversion, and demographics. *Journal of Economic Behavior & Organization*, *98*, 41-55.

Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological science*, *21*(4), 523-527.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child development*, *82*(1), 405-432.

Flood, A. M., Hare, D. J., & Wallis, P. (2011). An investigation into social information processing in young people with Asperger syndrome. *Autism*, *15*(5), 601-624.

Fontaine, R. G., & Dodge, K. A. (2006). Real-time decision making and aggressive behavior in youth: a heuristic model of response evaluation and decision (RED). *Aggressive Behavior*, *32*(6), 604-624.

Fudenberg, D., Rand, D. G., & Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, *102*(2), 720-49.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, *38*(5), 581-586.

Helseth, S. A., Waschbusch, D. A., King, S., & Willoughby, M. T. (2015). Aggression in children with conduct problems and callous-unemotional traits: social information processing and response to peer provocation. *Journal of abnormal child psychology*, 1-12.

Hubbard, J. A., McAuliffe, M. D., Morrow,M. T., & Romano, L. J. (2010). Reactive and proactive aggression in childhood and adolescence: Precursors, outcomes, processes, experiences, and measurement. Journal of Personality, 78, 95–118.

Hubbard, J. A., Smithmyer, C. M., Ramsden, S. R., Parker, E. H., Flanagan, K. D., Dearing, K. F., et al. (2002). Observational, physiological, and self-report measures

of children's anger: Relations to reactive versus proactive aggression. Child Development, 73, 1101–1118.

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of personality and social psychology*, *72*(2), 408.

Kempes, M., De Vries, H., Matthys, W., Van Engeland, H., & Van Hooff, J. (2008). Differences in cortisol response affect the distinction of observed reactive and proactive aggression in children with aggressive behaviour disorders. *Journal of Neural Transmission*, *115*(1), 139-147.

Kempes, M., Matthys, W., De Vries, H., & Van Engeland, H. (2005). Reactive and proactive aggression in children A review of theory, findings and the relevance for child and adolescent psychiatry. *European child & adolescent psychiatry*, *14*(1), 11-19.

Kempes, M., Matthys, W., Maassen, G., van Goozen, S., & van Engeland, H. (2006). A parent questionnaire for distinguishing between reactive and proactive aggression in children. *European child & adolescent psychiatry*, *15*(1), 38-45.

King-Casas, B., & Chiu, P. H. (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biological psychiatry*, *72*(2), 119-125.

Kopelman, S., Weber, J. M., & Messick, D. M. (2002). Factors influencing cooperation in commons dilemmas: A review of experimental psychological research. *The drama of the commons*, 113-156.

Kupersmidt, J. B., Stelter, R., & Dodge, K. A. (2011). Development and validation of the social information processing application: A Web-based measure of social information processing patterns in elementary school-age boys. *Psychological assessment*, *23*(4), 834.

Laible, D. J., Murphy, T. P., & Augustine, M. (2014). Adolescents' Aggressive and Prosocial Behaviors: Links With Social Information Processing, Negative Emotionality, Moral Affect, and Moral Cognition. *The Journal of genetic psychology*, *175*(3), 270-286.

Laible, D., McGinley, M., Carlo, G., Augustine, M., & Murphy, T. (2014). Does engaging in prosocial behavior make children see the world through rose-colored glasses?. *Developmental psychology*, *50*(3), 872.

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological science*, *17*(5), 421-427.

Maskin, E., & Fudenberg, D. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, *53*(3).

Matsumoto, D., Haan, N., Yabrove, G., Theodorou, P., & Carney, C. C. (1986). Preschoolers' moral actions and emotions in Prisoner's Dilemma.*Developmental Psychology*, *22*(5), 663.

Matthys, W., & Lochman, J. (2011). *Oppositional defiant disorder and conduct disorder in children*. John Wiley & Sons.

McClure-Tone, E. B., Nawa, N. E., Nelson, E. E., Detloff, A. M., Fromm, S. J., Pine, D. S., & Ernst, M. (2011). Preliminary findings: neural responses to feedback regarding

betrayal and cooperation in adolescent anxiety disorders. *Developmental neuropsychology*, *36*(4), 453-472.

McClure, E. B., Parrish, J. M., Nelson, E. E., Easter, J., Thorne, J. F., Rilling, J. K., ... & Pine, D. S. (2007). Responses to conflict and cooperation in adolescents with anxiety and mood disorders. *Journal of abnormal child psychology*, *35*(4), 567-577.

Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., & Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *Journal of abnormal psychology*, *117*(2), 406.

Muñoz, L. C., Frick, P. J., Kimonis, E. R., & Aucoin, K. J. (2008). Types of aggression, responsiveness to provocation, and callous-unemotional traits in detained adolescents. *Journal of abnormal child psychology*, *36*(1), 15-28.

Nelson, D. A., & Crick, N. R. (1999). Rose-colored glasses: Examining the social information-processing of prosocial young adolescents. *The Journal of Early Adolescence*, *19*(1), 17-38.

Normann, H. T., & Wallace, B. (2012). The impact of the termination rule on cooperation in a prisoner's dilemma experiment. *International Journal of Game Theory*, *41*(3), 707-718.

Perner, J. (1979). Young children's preoccupation with their own payoffs in strategic analysis of 2× 2 games. *Developmental Psychology*, *15*(2), 204.

Phillips, N. C., & Lochman, J. E. (2003). Experimentally manipulated change in children's proactive and reactive aggressive behavior. *Aggressive Behavior*, *29*(3), 215-227.

Polman, H., de Castro, B. O., Koops, W., van Boxtel, H. W., & Merk, W. W. (2007). A meta-analysis of the distinction between reactive and proactive aggression in children and adolescents. *Journal of Abnormal Child Psychology*, *35*(4), 522-535.

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., ... & Liu, J. (2006). The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive behavior*, *32*(2), 159.

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in cognitive sciences*, *17*(8), 413-425.

Rilling, J. K., Glenn, A. L., Jairam, M. R., Pagnoni, G., Goldsmith, D. R., Elfenbein, H. A., & Lilienfeld, S. O. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological psychiatry*, *61*(11), 1260-1271.

Sally, D., & Hill, E. (2006). The development of interpersonal strategy: Autism, theory-of-mind, cooperation and fairness. *Journal of economic psychology*, *27*(1), 73-97.

Sharp, C., Ha, C., & Fonagy, P. (2011). Get them before they get you: Trust, trustworthiness, and social cognition in boys with and without externalizing behavior problems. *Development and Psychopathology*, *23*(02), 647-658.

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology*, *28*(1), 69-78.

van Rest, M. M., van Bokhoven, I., van Nieuwenhuijzen, M., Embregts, P. J., Vriens, A., & Matthys, W. (2014). Developing a new assessment procedure of social

information processing in adolescents within secure residential care. *Research in developmental disabilities*, *35*(6), 1402-1411.

Vitaro, F., Brendgen, M., & Barker, E. D. (2006). Subtypes of aggressive behaviors: A developmental perspective. *International Journal of Behavioral Development*, *30*(1), 12-19.

Vogelsang, M., & Tomasello, M. (2016). Giving is nicer than taking: Preschoolers reciprocate based on the social intentions of the distributor. *PloS one*, *11*(1), e0147539.

Warneken, F., & Tomasello, M. (2013). The emergence of contingent reciprocity in young children. *Journal of experimental child psychology*, *116*(2), 338-350.

Waschbusch, D. A., Pelham, W. E., Jennings, J. R., Greiner, A. R., Tarter, R. E., & Moss, H. B. (2002). Reactive aggression in boys with disruptive behavior disorders: Behavior, physiology, and affect. *Journal of Abnormal Child Psychology*, *30*(6), 641-656.

Yaros, A., Lochman, J. E., Rosenbaum, J., & Jimenez-Camargo, L. A. (2014). Real-time hostile attribution measurement and aggression in children. *Aggressive behavior*, *40*(5), 409-420.